

Voor Simon

Promotoren Prof. dr. Gert De Sutter
Vakgroep Vertalen, Tolken en Communicatie
Prof. dr. Johan De Caluwe
Vakgroep Taalkunde

Decaan Prof. dr. Marc Boone
Rector Prof. dr. Anne De Paepe



Faculteit Letteren & Wijsbegeerte

Annelore Willems

Uit het veld geslagen

*Een corpusgebaseerd multivariaat onderzoek naar de
positie van voorzetselconstituenten in de Nederlandse
journalistieke bijzin*

Proefschrift voorgelegd tot het behalen van de graad van
Doctor in de Nederlandse taalkunde

2017

Dankwoord

Onderzoek is boeiend en uitdagend, maar zorgt tegelijkertijd wel eens voor gevoelens van onzekerheid. Voor mij en mijn onderzoek was dat zeker niet anders. De vraag ‘waarom onderzoek ik woordvolgordevariatie?’ is een vraag die ik me vaak heb gesteld, en die ik niet altijd even goed kon beantwoorden. In tijden van onzekerheid dacht ik aan een artikel uit 2011 over koolmezen, die variëren in hun gezang. Daarbij herinner ik me vooral de volgende zin: ‘Hoe meer variatie een koolmees in zijn liedje inbouwt, hoe groter zijn kans op overleven dat jaar’. Als variatie bij koolmezen zoveel teweeg kan brengen, wat zou variatie in de woordvolgorde van mensen dan wel niet kunnen bereiken? Deze kleine vogeltjes verdienen dan ook een prominente plaats in mijn proefschrift en sieren daarom de kaft.

Voor elk moeilijk moment tijdens dit proces zijn er minstens evenveel momenten van pure dankbaarheid. Het is daarom hoog tijd dat ik elk van jullie bedank en eerlijk, ik heb hier ontzettend naar uitgekeken.

Mijn grootste dankjewel verdien jij, Gert. Je bent mijn promotor, maar eigenlijk zoveel meer. Ik denk dat we al snel een vader-dochterrelatie hadden opgebouwd, of klinkt oudere broer-zus misschien beter voor jou? Na een tijdje kon je mij makkelijk lezen en je wist vaak al op voorhand wat ik zou antwoorden of zeggen. Je hielp me in elke fase van mijn assistentschap. Ik herinner me de try-out voor mijn eerste congres, de abstracts die je me liet schrijven en herschrijven, de vele gesprekken over het onderzoek en over mijn zorgen en verzuchtingen. Vooral jouw menselijkheid, je open blik en je professionele kwaliteiten wil ik in dit proefschrift uitdrukkelijk vermelden. Je hebt mij als de beste begeleid, in mijn onderzoek, maar even goed als lesgever en bovenal als persoon. Je hebt me zien groeien van studente tot mama en stond me altijd bij met wijze raad. Ik heb het geluk gehad jou als promotor te hebben en ik overdrijf niet als ik zeg dat er zonder jou nooit een proefschrift had gelegen (en dat is de waarheid, de enige waarheid).

Johan, ik herinner me vooral nog je lessen en mijn wens om bij jou een bachelor- en masterproef te schrijven. De sms-taal hebben we samen bestudeerd en op het einde van

de rit kreeg ik van jou de vraag of ik niet geïnteresseerd was in onderzoek. Ik zei ronduit ‘neen’, want ik wilde lesgeven. Pas toen je mij vertelde dat er ook zoiets als een assistentschap bestaat, waarbij onderzoek en onderwijs gecombineerd worden, wist je me te overtuigen. Door de jaren heen heb ik een andere Johan leren kennen. Een Johan die zijn ei kwijt kan in het brouwen van wijn, die als de beste kan tuinieren en die me altijd, samen met Annik, met een warm gevoel (inclusief kop koffie en zoetigheden) ontvangen heeft. Tijdens die momenten wist je me steeds aan te moedigen, ook al begreep je al snel dat mijn hart bij het lesgeven lag. Dankjewel voor je interesse in mijn onderzoek, in mijn leven naast het onderzoek en dankjewel dat ik je op een andere manier heb mogen leren kennen.

Ik wil ook graag de leden van mijn doctoraatsbegeleidingscommissie bedanken, Stefan Grondelaers en Timothy Coleman. Jullie hebben bijgedragen aan mijn doctoraat door mij te laten nadenken over operationalisaties en resultaten, door me te leren een kritische attitude aan te meten en door jullie inzichten, tijd en ervaring met mij te delen.

Een bijzondere dankjewel gaat ook uit naar mijn collega’s. Ik wil het aantal zuchten en de vele ‘waarom begint iemand aan een doctoraat?’ niet tellen. Jullie wisten mij telkens te motiveren tijdens de vele lezingen die we samen in ons kantoor hebben gehouden. Isabelle, we hebben elkaar in B1.18 leren kennen, toen zat ik recht tegenover jou, en je wist me al van bij het begin aan te moedigen. Zelfs in de laatste fase van mijn doctoraat stuurde je mij de juiste, stimulerende woorden en grappige foto’s (met ijsberen nota bene!). Ik sta na al die jaren nog altijd versteld van jouw organisatietalent en jouw aandacht voor de kleine en grote dingen in het leven. Koen, als er iemand de liefde voor statistiek in mij wist wakker te schudden, ben jij het wel. Je wist me zelfs over te halen meerdere statistiekcursussen te volgen, waarvoor ik je nog altijd dankbaar ben. Lore, we lijken niet alleen fysiek op elkaar, maar ook in vele andere aspecten vinden we elkaar. Ik denk daarbij spontaan aan onze liefde voor brocante en oude huizen, leuke (maar vaak veel te dure) designspulletjes voor onze zoontjes, lekkere (Chinese) thee en yoga. Ik bewonder jouw werkethos, jouw talenknobbel en jouw doorzettingsvermogen (voer maar eens een onderzoek uit waar je niet zelf voor gekozen hebt). Chloé, jij bent een werkpaard waar zelfs het grootste boerenpaard nog wat van kan leren. Ik geniet van jouw eerlijkheid, hoe je in het leven staat, jouw flair en natuurlijk van onze gesprekken. Tot slot nog een dankjewel aan onze twee jonkies, Lynn en Pauline. Jullie enthousiasme en jeugdigheid deed deugd in B1.16. Ik krijg spontaan een glimlach op mijn lippen als ik aan jouw to-do-lijstjes denk, Lynn, en ik kijk al meteen verlekkerd bij de gedachte dat je je aan een Meusje waagt. Dank je wel, Pauline, voor de babbels tijdens de lunch over verbouwingen en de zee. Het heeft mijn gedachten verzet wanneer ik het nodig had. Een zachte zeebries is nooit ver weg als jij er bent.

Stefaan, dank je wel om altijd tijd te maken, mij te motiveren als onderzoeker en ruimte te geven (en te begeleiden!) als lesgever. Ook de collega's van B1.18 wil ik hierbij graag bedanken. Rita, ik heb het geluk gehad jou het laatste academiejaar te mogen volgen. Jouw ontzettend uitgebreide woordenschat, jouw enthousiasme en jouw uitstraling als lesgever dagen ook mij uit om de beste lesgever in mezelf naar boven te halen. Annik, jij bent waarschijnlijk de liefste persoon die ik ken. Ik waardeer je ijver voor onze studenten en je wilskracht om jouw zonen een warme thuis te geven. Het is altijd een plezier om even bij jou langs te komen, op het werk of in Latem. Filip, bij elke grammaticavraag kon ik een beroep doen op jou. Je hebt me er meermaals op gewezen dat grammatica geen zwart-witverhaal is, om me vervolgens te sussen dat het allemaal wel goed komt. Ruud, jou heb ik vooral op weg naar het sollicitatieseminarie leren kennen (ik herinner me nog altijd jouw ervaring in de Joodse school) en het voorbije jaar tijdens de mondelinge examens. Luisteren naar jou en naar jouw prachtige stem, is altijd een plezier. Sylvianne, mijn maatje, je was er de voorbije jaren altijd, met een lach, een bemoedigend woord, een schouderklop of een goed gesprek. Elke keer verlichtte dat mijn pas en was het net dat ietsje makkelijker om door te zetten. Tot slot wil ik in het bijzonder Myriam bedanken. Je was (en bent) mijn mentor die me van in het begin heeft begeleid. Ik onthoud jouw norske blik (niet mijn woorden, maar die van jou), die ik met momenten zelf graag uitprobeer, maar vooral jouw enthousiasme en jouw energie om studenten warm te maken Standaardnederlands te spreken.

Naast de collega's van de afdeling Nederlands met wie ik de voorbije jaren heb samengewerkt, zijn er uiteraard nog collega's die ik graag wil bedanken. Mijn eerste dank gaat uit naar Veronique. Jouw deur staat altijd open, voor grote of kleine vragen. Ik bewonder jouw werk – als vakgroepvoorzitter, onderzoeker en als lesgever – jouw vertrouwen in mij en jouw warme lach tijdens de lunch in de docentenkamer. Ook het administratieve personeel verdient een grote dankjewel. Ann, Sofie, Ann, Chantal, Nadine en Linda, jullie werk achter de schermen is onbeschrijfelijk en ook de laatste weken, voor de uiteindelijke verdediging, hebben jullie getoond dat jullie werkelijk bergen kunnen verzetten. Tot slot wil ik ook de vele andere collega's bedanken: Orphée, mijn collega en buurman Bart, Bernard, Sarah, Sabine, Griet, ... Bedankt voor de heerlijke babbels, voor het plezier in de gang en de docentenkamer, voor de personen die jullie zijn.

Lieve vrienden, jullie verdienen uiteraard ook een vermelding in dit dankwoord. De laatste maanden was het vaak stil aan mijn kant. Ondanks die stilte vergeten jullie mij niet en blijven jullie berichtjes sturen om te vragen hoe het gaat. Valerie, Tina, Cilia, Julia, ik kan zo ontzettend genieten van onze lunches, kerstfeestjes en skype dates. Herinneringen ophalen met jullie doet me altijd even stilstaan bij onze jarenlange vriendschap. Jullie doen me een voor een stralen! Sarah, de grappigste en meest romantische films zag ik samen met jou. En laten we de vele ribbikes niet vergeten,

uiteraard met twee (of drie?) patatten. Ine en Lieselot, jullie eerlijkheid en kijk op het leven maken het altijd een plezier om af te spreken. Laurence en Thomas, de vele weekendjes zee, de etentjes en zelfs ons weekendje zonder kindjes, hebben me de rust gegeven die ik met momenten zo nodig had. Aan iedereen: bedankt om er al die jaren voor mij te zijn. Jullie verrijken mijn leven en ik noem mezelf een echte gelukzak met zulke vrienden.

Uiteraard wil ik ook heel graag mijn familie en schoonfamilie bedanken voor al hun aanmoedigingen en hun onwrikbare geloof in mij. Jullie hebben me gevormd tot wie ik ben. Geneviève, dank je wel voor jouw gastvrijheid elk weekend en voor het lekkerste eten ooit. Ghislain, ik blijf in bewondering voor je ongelooflijke dorst naar kennis. Tom, je bent na al die jaren de broer die ik nooit heb gehad. Het blijft een plezier om samen met jou Matthias op stang te jagen met mijn betere rijcapaciteiten (wat Matthias natuurlijk zal tegenspreken). Annelien, eind 2016 diende jij jouw doctoraat in. Jouw gedrevenheid stimuleerde me misschien wel nog meer om mijn proefschrift ook af te werken en in te dienen. In het bijzonder wil ik mijn papa bedanken. Als klein meisje was je mijn grote voorbeeld en dat ben je nog altijd. Ik kijk met veel bewondering en respect naar de opvoeding die ik gekregen heb. Jij hebt me het nodige vertrouwen gegeven om zelf mijn pad te kiezen en je hebt me geleerd door te zetten als het wat moeilijker gaat. Je hebt urenlang mijn lessen overhoord. Zelfs toen ik de pagina's volledig gememoriseerd had en jij het wel welletjes vond, bleef je toch doorgaan tot ik zelf geloofde dat ik de stof beheerste. Soms kon ik razend worden omdat je mijn gestress minimaliseerde - want "Loortje, het komt wel goed" - maar uiteindelijk gaf me dat altijd dat extra duwtje in de rug. Onze band is uniek en dat zal iedereen in onze omgeving beamen. Dank je wel, papa, voor je vertrouwen, je hulp in elke levensfase en je onvoorwaardelijke steun.

Mijn allergrootste liefde gaat naar jou, Matthias. Op elk moment in mijn doctoraatsonderzoek heb je me geholpen, gaande van teksten nalezen, nadenken over titels en lay-out tot zelfs Engelse presentaties inspreken. Je laat me razen en briesen als ik het nodig heb, om mij vervolgens met de juiste woorden op te vangen. Ik besef dat mijn proefschrift ook voor jou soms een uitdaging was: want hoe zorg je ervoor dat een doctoraatsstudente haar hoofd koel houdt, laat staan een zwangere? Maar Matthias, je hebt nooit geklaagd en je hield van me tijdens het volledige proces. Dankjewel om mij altijd weer met de voeten op de grond te brengen en om me even een spiegel voor te houden. Dankjewel om er voor mij te zijn en om mij gelukkig te maken. Voor jou - meer dan voor wie ook - ben ik blij dat mijn proefschrift er ligt

Lijst afkortingen

AdjC	Adjectiefconstituent of Adjectivische Constituent
AdvC	Adverbiale Constituent
AIC	Akaike Information Criterion
ANOVA	Analysis of Variance
ANS	Algemene Nederlandse Spraakkunst
AV	Achterveld
CP	Complementizer Phrase
DLM	Dependency Length Minimization
DPC	Dutch Parallel Corpus
GLM	Generalized Linear Model
Log2	Logaritme met grondtal 2
MV	Middenveld
NA	Not Available
NC of NP	Nominale of Naamwoordelijke Constituent of Noun Phrase
LF	Logical Form
LRM	Logistic Regression Model
PCA	Principal Component Analysis
PC	Principal Component
PF	Phonetic Form
SOV	Subject – Object – Verb
SVO	Subject – Verb – Object
SVx	Subject – Verb – x (= constituent)
SxV	Subject – x (= constituent) – Verb
Vif	Variance inflation factor
VzAzC	Voorzetsel-achterzetselconstituent
VzC of PP	Voorzetselconstituent of Prepositional Phrase
Vgw	Voegwoord
VP	Verb Phrase
Vz of P	Voorzetsel of Preposition
Ww of V	Werkwoord of Verb

Lijst tabellen

Tabel 1	Voorstelling Gibsons theorie (uit Jansen 2003: 88).....	55
Tabel 2	Voorstelling Gibsons theorie bij een zin met twee ingevulde polen	56
Tabel 3	Voorstelling Gibsons theorie met aanpassing van Jansen (2003: 90).....	56
Tabel 4	Overzicht van de onderzochte variabelen in empirisch onderzoek (§2.2.4)	59
Tabel 5	Overzicht variabelen op basis van het bestaande empirische onderzoek (§2.2.4) met hun voorkeur voor middenveld- of achterveldpositie	59
Tabel 6	Overzicht journalistieke teksten in het DPC (uit Delaere 2015).....	62
Tabel 7	Overzicht aantal woorden en lettergrepen voor het middenveld, achterveld en de VzC.....	82
Tabel 8	Overzicht van de determinatoren die bepalen of de VzC <i>definit</i> of <i>indefinit</i> is (gebaseerd op Coussé 2009).....	86
Tabel 9	Variabele <i>toegankelijkheid</i> : mathematische waarde (kolom 1); labels (kolom 2), rangorde (kolom 3); labels die significant van elkaar verschillen na gepartitioneerde chikwadraattoets (kolom 4); rangorde na gepartitioneerde chikwadraattoets (kolom 5).....	91
Tabel 10	Grammaticale functies van de VzC en een nieuwe onderverdeling met behulp van de gepartitioneerde chikwadraattoets	96
Tabel 11	Semantische categorieën van de VzC en een nieuwe onderverdeling met behulp van de gepartitioneerde chikwadraattoets	97
Tabel 12	Overzicht van de 10 meest voorkomende werkwoorden in onze studie en in het SUBTLEX-NL-corpus	101
Tabel 13	Semantische onderverdeling werkwoorden volgens Referentiebestand Nederlands	101
Tabel 14	Kolom 1 en 2 is de subclassificatie volgens Maks, van der Vliet, Görög & Vossen (2013: 12-13); kolom 3 onze eigen semantische (sub)classificatie.....	103
Tabel 15	De verdeling van het werkwoord <i>betalen</i> tegenover alle andere werkwoorden.....	104
Tabel 16	Werkwoorden die op basis van de collostructional analysis (Gries & Stefanowitsch 2004) een voorkeur hebben voor een van de beide woordvolgordes ($p < 0.05$).....	104
Tabel 17	Onderverdeling variabele <i>zinsrelator</i>	105

Tabel 18	Shapiro-Wilk-test voor de verschillende operationalisaties van <i>syntactische complexiteit</i> middenveld116
Tabel 19	Zin 101 ontleed; hiërarchie (of de niveaus) in kolom 1 en het bijzintype in kolom 3.....120
Tabel 20	De relatie tussen <i>lengte in woorden</i> en <i>lengte in lettergrepen</i> : overzicht van de standaardafwijking, de variantie en de cumulatieve variantie van de 6 hoofdcomponenten128
Tabel 21	De relatie tussen <i>lengte in woorden</i> en <i>inbedding</i> : overzicht van de standaardafwijking, de variantie en de cumulatieve variantie van de 6 hoofdcomponenten131
Tabel 22	De relatie tussen <i>het topologische middenveld</i> en <i>het cognitief congruerende middenveld</i> : overzicht van de standaardafwijking, de variantie en de cumulatieve variantie van de 2 hoofdcomponenten133
Tabel 23	Overzicht absolute frequentie per geattesteerde semantische classificatie.....147
Tabel 24	Vergelijking variabelen <i>type hoofdwerkwoord</i> en <i>zinsrelator</i>149
Tabel 25	Overzicht predictorvariabelen: hoofdeffecten en interactie-effecten153
Tabel 26	Logistische regressie van de toename van achterveldpositie tegenover middenveldpositie voor het journalistieke genre van het DPC corpus (n = 1711, AIC = 1798,3)155
Tabel 27	Logistische regressie van de toename van achterveldpositie tegenover middenveldpositie voor het journalistieke genre van het DPC corpus na eliminatie van de niet-significante interactie-effecten (n = 1711, AIC = 1780,1)156
Tabel 28	De minimale logistische regressie van de toename van achterveldpositie tegenover middenveldpositie voor het journalistieke genre van het DPC corpus (n = 1711, AIC = 1765,9)157
Tabel 29	<i>Variance inflation factor</i> van het uiteindelijke model161
Tabel 30	<i>Variance inflation factor</i> van het uiteindelijke model model zonder interactie-effecten161
Tabel 31	Overzicht probabiliteit en 95%-betrouwbaarheidsinterval voor de variabele <i>lengte van het middenveld</i>164
Tabel 32	Overzicht probabiliteit en 95%-betrouwbaarheidsinterval voor de variabele <i>semantische status van de VzC</i>165
Tabel 33	Overzicht probabiliteit en 95%-betrouwbaarheidsinterval voor de variabele <i>type hoofdwerkwoord</i>167
Tabel 34	Overzicht probabiliteit en 95%-betrouwbaarheidsinterval voor de variabele <i>zinsrelator</i>168
Tabel 35	Overzicht probabiliteit en 95%-betrouwbaarheidsinterval voor de interactie <i>Lengte_AV</i> en <i>Lengte_VzC</i>170
Tabel 36	Overzicht probabiliteit en 95%-betrouwbaarheidsinterval voor de interactie <i>Toegankelijkheid</i> en <i>Belang_na</i>172

Lijst figuren

Figuur 1	T-model (uit Hoekstra 1984).....	31
Figuur 2	Overzicht Case-marked positions (Hoekstra 1984: 82).....	33
Figuur 3	Verplichte linkse verplaatsing van complementen volgens Koster (2001b)	37
Figuur 4	Parallele specificatie van het target [e] volgens Koster (2001b).....	38
Figuur 5	Syntaxis volgens Barbiers VP-Intrapositie (Barbiers 2008: 172).....	41
Figuur 6	Schematisch voorstelling van het eerste selectie criterium: <i>de afstand tussen het bindwoord en het vervoegde werkwoord is maximaal 16 woorden</i>	64
Figuur 7	Beslissingsboom zinsdeel versus zinsdeelstuk	68
Figuur 8	Staafdiagram van de verdeling middenveld- en achterveldpositie per journalistiek subgenre	77
Figuur 9	Staafdiagram van de verdeling middenveld- en achterveldpositie in niet-vertaald en vertaald Nederlands.....	79
Figuur 10	Taxonomie van de variabele <i>toegankelijkheid</i> (gebaseerd op Karreman, Enschoot & Schuur 2013).....	90
Figuur 11	Staafdiagram van de afhankelijke variabele: overzicht van de relatieve frequentie (y-as) middenveldpositie en achterveldpositie (x-as)	112
Figuur 12	Histogram (links) en boxplot (rechts) van de <i>lengte in woorden van het middenveld</i>	114
Figuur 13	Histogram (links) en boxplot (rechts) van de <i>lengte in lettergrepen van het middenveld</i>	114
Figuur 14	Histogram (links) en boxplot (rechts) van de <i>inbedding van het middenveld</i>	115
Figuur 15	Histogram (links) en boxplot (rechts) van de log2-transformatie van het <i>aantal woorden in het middenveld</i>	117
Figuur 16	Histogram (links) en boxplot (rechts) van de log2-transformatie van het <i>aantal lettergrepen in het middenveld</i>	117
Figuur 17	Histogram (links) en boxplot(rechts) van de log2-transformatie van het <i>aantal inbeddingen in het middenveld</i>	117
Figuur 18	Histogram (links) en boxplot (rechts) van de <i>lengte in woorden van het achterveld</i>	119
Figuur 19	Histogram (links) en boxplot (rechts) van de <i>lengte in lettergrepen van het achterveld</i>	119
Figuur 20	Histogram (links) en boxplot (rechts) van de <i>inbedding van het achterveld</i>	120

Figuur 21	Histogram (links) en boxplot (rechts) van de log2-transformatie van het <i>aantal woorden in het achterveld</i>121
Figuur 22	Histogram (links) en boxplot (rechts) van de log2-transformatie van het <i>aantal lettergrepen in het achterveld</i>121
Figuur 23	Histogram (links) en boxplot (rechts) van de log2-transformatie van het <i>aantal inbeddingen in het achterveld</i>121
Figuur 24	Histogram (links) en boxplot (rechts) van de <i>lengte in woorden van de VzC</i>122
Figuur 25	Histogram (links) en boxplot (rechts) van de <i>lengte in lettergrepen van de VzC</i>123
Figuur 26	Histogram (links) en boxplot (rechts) van het <i>aantal inbeddingen van de VzC</i>123
Figuur 27	Histogram (links) en boxplot (rechts) van de log2-transformatie van het <i>aantal woorden in de VzC</i>124
Figuur 28	Histogram (links) en boxplot (rechts) van de log2-transformatie van het <i>aantal lettergrepen in de VzC</i>124
Figuur 29	Histogram (links) en boxplot (rechts) van de log2-transformatie van het <i>aantal inbeddingen in de VzC</i>124
Figuur 30	Overzicht en interpretatie van de correlatiecoëfficiënt (Gries 2013: 147)126
Figuur 31	Grafische voorstelling van de correlerende operationalisaties van <i>lengte (in woorden en lettergrepen) van het middenveld, achterveld en de VzC</i>127
Figuur 32	De relatie tussen <i>lengte in woorden</i> en <i>lengte in lettergrepen</i> : de relatieve variantie (y-as) per PC (x-as)128
Figuur 33	Correlatiecirkel van de <i>lengte in woorden (log2)</i> en de <i>lengte in lettergrepen (log2)</i> op basis van PC1 (x-as) en PC2 (y-as)129
Figuur 34	Grafische voorstelling van <i>inbedding</i> en <i>lengte in woorden van het middenveld, achterveld en de VzC</i>129
Figuur 35	De relatie tussen <i>lengte in woorden</i> en <i>inbedding</i> : de relatieve variantie (y-as) per PC (x-as)130
Figuur 36	Correlatiecirkel van de <i>lengte in woorden (log2)</i> en het <i>aantal inbeddingen (log2)</i> op basis van PC1 (x-as) en PC2 (y-as)131
Figuur 37	Spreidingdiagram <i>lengte (in woorden) topologisch en cognitief middenveld</i>133
Figuur 38	De relatie tussen <i>het topologische middenveld</i> en <i>het cognitief congruerende middenveld</i> : de relatieve variantie (y-as) per PC (x-as)133
Figuur 39	Correlatiecirkel van het <i>topologische middenveld (log2)</i> en het <i>cognitief congruerende middenveld (log2)</i> op basis van PC1 (x-as) en PC2 (y-as)134
Figuur 40	Staafdiagram van de variabele <i>definietheid</i> : overzicht van de relatieve frequentie (y-as) definiete en indefiniete VzC's (x-as)135
Figuur 41	Staafdiagram van de variabele <i>toegankelijkheid</i> (8 categorieën): overzicht van de relatieve frequentie (y-as) per categorie (x-as)136
Figuur 42	Histogram (links) en boxplot (rechts) van <i>informationeel belang: voorcontext</i>138
Figuur 43	Histogram (links) en boxplot (rechts) van de log2-transformatie van <i>informationeel belang: voorcontext</i>138

Figuur 44	Histogram (links) en boxplot (rechts) van <i>informationeel belang: nacontext</i>	139
Figuur 45	Histogram (links) en boxplot (rechts) van de log2-transformatie van <i>informationeel belang: nacontext</i>	139
Figuur 46	Gecombineerd staafdiagram van de variabele <i>definietheid (def/indef)</i> per level van de variabele <i>toegankelijkheid</i>	140
Figuur 47	Staafdiagram <i>grammaticale status van de VzC</i>	142
Figuur 48	Staafdiagram <i>semantische status van de VzC</i>	143
Figuur 49	Histogram (links) en boxplot (rechts) van <i>frequentie van het hoofdwerkwoord</i>	144
Figuur 50	Histogram (links) en boxplot (rechts) van de <i>frequentie van het hoofdwerkwoord</i> (zonder de werkwoorden <i>hebben, zijn</i>)	145
Figuur 51	Histogram (links) en boxplot (rechts) van de log2-transformatie van de variabele <i>frequentie</i>	146
Figuur 52	Staafdiagram van de variabele <i>semantische classificatie van het werkwoord</i>	146
Figuur 53	Staafdiagram van de variabele <i>type hoofdwerkwoord</i>	147
Figuur 54	Staafdiagram van de variabele <i>zinsrelator</i>	148
Figuur 55	Output R van de uiteindelijke minimale logistische regressie met behulp van GLM	157
Figuur 56	Output R van de uiteindelijke minimale logistische regressie met behulp van LRM.....	158
Figuur 57	Probabiliteitsschaal	162
Figuur 58	Effect plot van <i>lengte middenveld</i>	163
Figuur 59	Effect plot van de <i>semantische classificatie van de VzC</i>	165
Figuur 60	Effect plot voor <i>type hoofdwerkwoord</i>	166
Figuur 61	Effect plot van de <i>zinsrelator</i>	168
Figuur 62	Effect plot van de interactie <i>lengte van het AV (log2)</i> en <i>lengte van de VzC (log2)</i>	169
Figuur 63	Effect plot van de interactie <i>toegankelijkheid</i> en <i>belang nacontext (log2)</i>	171
Figuur 64	Calibratiecurve met 1000 bootstrapherhalingen	174
Figuur 65	Model van het cognitieve schrijfproces volgens Flower & Hayes (1981: 370)	181

Inhoudstafel

Hoofdstuk 1	Inleiding.....	1
Hoofdstuk 2	Literatuurstudie	9
2.1	Algemene literatuur	10
2.1.1	Grammatica's	10
2.1.2	Stijlgidsen	16
2.1.3	Conclusie algemene literatuur	18
2.2	Wetenschappelijke literatuur	20
2.2.1	Historisch onderzoek.....	20
2.2.1.1	Diachroon overzicht van de zinsdelen in achterveldpositie.....	20
2.2.1.2	Variabelen die achterveldpositie verklaren.....	22
2.2.1.3	Conclusie historisch onderzoek	27
2.2.2	Generatief onderzoek	27
2.2.2.1	De constructiespecifieke opvatting	30
2.2.2.2	De principes-en-parameters-visie.....	31
2.2.2.3	De minimalistische opvatting	35
2.2.2.4	Conclusie generatief onderzoek	41
2.2.3	Functioneel grammaticaal onderzoek.....	42
2.2.3.1	Functionele (discourse) grammatica	42
2.2.3.2	Constructiegrammatica	44
2.2.4	Empirisch onderzoek	46
2.2.4.1	Eerste onderzoeken	46
2.2.4.1.1	Jansen (1978, 1979).....	46
2.2.4.1.2	Braecke (1990)	51
2.2.4.2	Psycholinguïstisch geïnspireerd onderzoek: Jansen (2003)	53
2.2.4.3	Corpusgebaseerd onderzoek naar de positie van voorzetselvoorwerpen: Hoeksema (2014).....	57
2.2.4.4	Conclusie empirische studies.....	59
Hoofdstuk 3	Methodologie	61
3.1	Corpus en datasetselectie	62
3.1.1	Dutch Parallel Corpus	62
3.1.2	Selectiecriteria.....	63
3.1.2.1	De afstand tussen het bindwoord en het vervoegde werkwoord is maximaal 16 woorden.....	63

3.1.2.2	De VzC komt voor in een bijzin ingeleid door een grammaticaal onderschikkend voegwoord <i>dat</i>	66
3.1.2.3	De VzC moet een zinsdeel zijn, geen zinsdeelstuk	66
3.1.2.4	De VzC kan zowel in het middenveld als in het achterveld staan	71
3.1.2.5	De VzC wordt niet voorafgegaan door een focuspartikel en is geen voornaamwoordelijk bijwoord, geen beknopte bijzin, geen <i>aan het + infinitief</i> -constructie	74
3.1.3	Het effect van de journalistieke subgenres	76
3.1.4	Het effect van vertaald en niet-vertaald Nederlands	78
3.2	Variabelen.....	80
3.2.1	De syntactische complexiteit	80
3.2.1.1	Toelichting bij de operationalisatie op basis van de <i>lengte</i>	81
3.2.1.2	Toelichting bij de operationalisatie op basis van de <i>inbedding</i>	82
3.2.1.3	Het topologische middenveld versus het cognitief congruerende middenveld	83
3.2.1.4	Hypotheses	84
3.2.2	De discoursstatus van de VzC	85
3.2.2.1	Toelichting bij de variabele <i>definietheid</i>	86
3.2.2.2	Toelichting bij de variabele <i>toegankelijkheid</i>	87
3.2.2.3	Toelichting bij de variabele <i>informationeel belang</i>	91
3.2.2.4	Hypotheses	94
3.2.3	De grammaticale en semantische status van de VzC's	95
3.2.3.1	Hypotheses	97
3.2.4	Het werkwoord	99
3.2.4.1	Toelichting bij de variabele <i>frequentie van het hoofdwerkwoord</i>	100
3.2.4.2	Toelichting bij de <i>semantische classificatie van het werkwoord</i> op basis van het Referentiebestand Nederlands	101
3.2.4.3	Toelichting bij de Collostructional Analysis (Gries & Stefanowitsch 2004): de variabele <i>type hoofdwerkwoord</i> en <i>zinsrelator</i>	103
3.2.4.4	Hypotheses	105
Hoofdstuk 4	Resultaten	109
4.1	Een descriptieve exploratie van de verklarende variabelen.....	111
4.1.1	De positie van de VzC.....	112
4.1.2	De syntactische complexiteit	113
4.1.2.1	Syntactische complexiteit van het middenveld: <i>lengte</i> en <i>inbedding</i>	113
4.1.2.2	Syntactische complexiteit van het achterveld: <i>lengte</i> en <i>inbedding</i>	118
4.1.2.3	Syntactische complexiteit van de VzC: <i>lengte</i> en <i>inbedding</i>	122
4.1.2.4	De relatie tussen de complexiteitsoperationalisaties: correlatietest en Principal Component Analysis	125
4.1.2.4.1	De relatie tussen <i>lengte</i> in woorden en <i>lengte</i> in lettergrepen	126

4.1.2.4.2	De relatie tussen lengte in woorden en inbedding	129
4.1.2.4.3	Samenvatting van de relatie tussen de complexiteitsoperationalisaties	132
4.1.2.5	De relatie tussen het topologische middenveld en het cognitief congruerende middenveld	132
4.1.3	Discoursstatus	134
4.1.3.1	De discoursstatus op basis van <i>definietheid</i>	135
4.1.3.2	De discoursstatus op basis van <i>toegankelijkheid</i>	136
4.1.3.3	De discoursstatus op basis van <i>informationeel belang</i>	137
4.1.3.3.1	Operationalisatie van <i>informationeel belang</i> : voorcontext.....	137
4.1.3.3.2	Operationalisatie van <i>informationeel belang</i> : nacontext	138
4.1.3.4	De relatie tussen <i>definietheid</i> en <i>toegankelijkheid</i>	139
4.1.4	De grammaticale en semantische status van VzC's.....	142
4.1.4.1	De <i>grammaticale status</i> van VzC's	142
4.1.4.2	De <i>semantische status</i> van VzC's.....	143
4.1.5	Het werkwoord	144
4.1.5.1	De onderverdeling van het werkwoord op basis van de <i>frequentie</i>	144
4.1.5.2	De onderverdeling van het werkwoord op basis van de <i>semantische classificatie</i>	146
4.1.5.3	De onderverdeling van het werkwoord naar het <i>type hoofdwerkwoord</i>	147
4.1.5.4	De onderverdeling van het werkwoord naar de <i>zinsrelator</i>	148
4.1.5.5	De relatie tussen het <i>type hoofdwerkwoord</i> en de <i>zinsrelator</i>	148
4.2	Een multivariate analyse van de verklarende variabelen	150
4.2.1	Logistische regressie	150
4.2.1.1	Van het maximale model naar het uiteindelijke model	154
4.2.2	Multicollineariteit	160
4.2.3	De bespreking en visualisatie van de significante predictoren.....	162
4.2.3.1	Hoofdeffect 1: de <i>lengte van het middenveld</i>	163
4.2.3.2	Hoofdeffect 2: de <i>semantische status van de VzC</i>	164
4.2.3.3	Hoofdeffect 3: het <i>type hoofdwerkwoord</i>	166
4.2.3.4	Hoofdeffect 4: de <i>zinsrelator</i>	167
4.2.3.5	Interactie-effect 1: de <i>lengte van het achterveld</i> en de <i>lengte van de VzC</i>	168
4.2.3.6	Interactie-effect 2: <i>toegankelijkheid</i> en <i>informationeel belang_na</i>	170
4.2.4	Verklarende en voorspellende kracht van minimale model	172
4.3	Besluit	175
Hoofdstuk 5 Discussie en conclusie.....		177
5.1	De VzC's verkiezen achterveldpositie	179
5.2	De predictoren op een functioneel-cognitieve manier verklaard	183
5.2.1	De semantische status van de VzC	183

5.2.2	De syntactische complexiteit	186
5.2.3	De werkwoordsvariabelen: type hoofdwerkwoord en zinsrelator	191
5.2.4	De discoursstatus	196
5.3	Vervolgonderzoek	198
Bibliografie	201
Appendix	211
Samenvatting	215
Summary	218

Hoofdstuk 1

Inleiding

De volgorde van de woorden binnen een zin is voor de meeste taalgebruikers een onopvallend aspect van de taal. De woorden zelf vallen meer op dan de volgorde waarin ze staan. Toch is die volgorde in het geheel van de taal zeer belangrijk. Zodra iemand een afwijkende woordvolgorde gebruikt, daarom niet noodzakelijk een foutieve, valt dat ineens wel op. Het nogal gecompliceerde systeem van woordvolgorderegels dat we dagelijks onbewust en als vanzelfsprekend hanteren, wordt dan zichtbaar.

Hoewel de meeste taalgebruikers de grammatica als een set van vaste syntactische regels zien, is de schikking van woorden zelfs in talen met een relatief vaste woordvolgorde aan variatie onderhevig. Variatiefenomenen die bijvoorbeeld in de neerlandistische literatuur meermaals belicht werden, zijn de positie van een direct object tegenover een bijwoord (ook wel *scrambling*; *Sonja heeft gisteren de kaas opgegeten* vs. *Sonja heeft de kaas gisteren opgegeten*; o.a., van Bergen & de Swart 2010), de positie van een indirect object (ook wel datiefalternantie; *Hij geeft het boek aan Marie* vs. *Hij geeft Marie het boek*; o.a., Coleman 2010) en de volgorde van werkwoorden binnen de werkwoordelijke eindgroep (ook wel de groene en rode volgorde: *De afstand die door mij gelopen is* vs. *De afstand die door mij is gelopen*; o.a., De Sutter 2007; Coussé 2008; Coussé et al. 2008). Vanuit verschillende achtergronden, met verschillende doelstellingen en methodes hebben taalwetenschappers zich over dergelijke en andere fenomenen van syntactische variatie gebogen. Verschillende aspecten werden daarbij belicht die onder meer voor belangrijke inzichten in de discursieve functies van syntactische variatie zorgden. Het Language-Independent Preferred Order of Constituents van Dik (1978), bijvoorbeeld, stelt vast dat constituenten meer naar achteren in de zin staan naarmate ze complexer (en dus langer) zijn. Het Code-Quantity Principle van Givón (1983) nuanceert het voorgaande en zegt dat

complexiteit ondergeschikt is aan de discoursstatus van constituenten¹. Naast deze discursieve functies werpt onderzoek naar syntactische variatie ook een licht op onderliggende cognitieve processen. Hawkins Principle of Early Immediate Constituents (1994) en Gibsons Dependency Locality Theory (2000) demonstreren allebei hoe het uitstellen van complexe constituenten tot het einde van de zin ervoor zorgt dat de hoorder sneller de overkoepelende structuur van de zin kan bepalen, waardoor de taalgebruiker minder moet onthouden en er een grote cognitieve belasting vermeden wordt.

In deze studie willen we een relatief onontgonnen type van syntactische variatie belichten, meer bepaald de variabele positie van voorzetselconstituenten (afgekort VzC's) in de Nederlandse bijzin. In het volgende voorbeeld zijn de twee woordvolgordes (1a) *deelnamen aan de Olympische Spelen* (VzC na de werkwoordelijke eindgroep; achterveldpositie) en (1b) *aan de Olympische Spelen deelnamen* (VzC voor de werkwoordelijke eindgroep; middenveldpositie) grammaticaal equivalent.

- (1) a [...] dat de Belgen deelnamen **aan de Olympische Spelen**
 b [...] dat de Belgen **aan de Olympische Spelen** deelnamen

Om de plaatsingsmogelijkheden van VzC's goed te begrijpen, is het nodig even kort in te gaan op het positionele zinspatroon van de Nederlandse hoofd- en bijzin.

Een karakteristieke eigenschap van het Nederlands is dat een zin uit twee vaste (verbale) polen bestaat. Omdat die polen meestal niet direct op elkaar volgen, verdelen ze de hoofdzin in drie zogenaamde topologische velden: een veld voor de eerste pool (het voorveld), een tussen de eerste en tweede pool (het middenveld) en een na de tweede pool (het achterveld)². De twee vaste polen vormen een denkbeeldige tang rond alle zinselementen in het middenveld, vandaar ook de benaming tangconstructie (o.a., Van Haeringen 1947; Haeseryn et al. 1997: 1225; Zwart 2011: 31-79).

In zin 2 (voorbeeld uit de ANS 1997: 1372) staat het vervoegde (hulp)werkwoord *kun* in de eerste pool en het hoofdwkwoord *zeggen* in de twee pool. Het direct object *zoiets* staat in het voorveld, het subject *je* en de zinsmodificeerder *wel* in het middenveld. Het voorzetselvoorwerp *tegen Kaatje* staat tot slot in het achterveld.

¹ Korte constituenten corresponderen met informatie die uit de voorgaande context bekend is en waar de spreker weinig aandacht op wil vestigen. Constituenten met meer fonologisch gewicht drukken contextueel minder voorspelbare of minder toegankelijke informatie uit die verdere explicitering nodig heeft (Givón 1988: 249). De plaatsing van korte, niet-complexe constituenten voor langere, complexere constituenten weerspiegelt m.a.w. de informatieve geleding van de zin. Eerst wordt het topic geïntroduceerd (of thema, het uitgangspunt van de zin) en vervolgens de comment of wat de taalgebruiker daarover te zeggen heeft (ook wel rhema, vaak het relatief belangrijkste gedeelte van een zin).

² We merken hierbij op dat elk veld leeg kan blijven en dat in de hoofdzin de tweede pool niet ingevuld kan zijn (Haeseryn et al. 1997: 1226).

- (2) Zoiets [kun]_{Pool1} je wel [zeggen]_{Pool2} **tegen Kaatje**.

Vrijwel alle zinsdelen in de vorm van een VzC kunnen in de drie beschikbare velden van een hoofdzin voorkomen. De VzC *tegen Kaatje* kan dus evengoed in het voorveld (voor de eerste pool; zie zin 2a) of in het middenveld (tussen de eerste en tweede pool; zie zin 2b) staan:

- (2) a **Tegen Kaatje** [kun]_{Pool1} je zoiets wel [zeggen]_{Pool2}.
b Zoiets [kun]_{Pool1} je **tegen Kaatje** wel [zeggen]_{Pool2}.

De topologie van Nederlandse bijzinnen verschilt van die van Nederlandse hoofdzinnen. Vergelijken we zin 2 (hoofdzin) met zin 3 (bijzin) dan valt het op dat de eerste pool in de bijzin een bindwoord³ is (in dit geval een grammaticaal onderschikkend voegwoord *dat*) en dat de tweede pool alle werkwoorden bevat (*kunt zeggen*). De tweede pool in een bijzin bestaat bijgevolg uit minstens één werkwoord, wat niet altijd het geval is in hoofdzinnen. Doordat de tweede pool in bijzinnen nooit leeg is, is er een zichtbare scheiding tussen de topologische velden midden- en achterveld. Een ander verschil is dat de bijzin geen voorveld heeft (ANS 1997: 1230, Zwart 2011: 35). De variabele positie van VzC's is in bijzinnen dus dichotoom van aard, want de mogelijke posities beperken zich tot twee velden: het middenveld (zie zin 3a) en het achterveld (zie zin 3b). De duidelijke opsplitsing tussen de topologische velden en het feit dat er maar twee mogelijke posities zijn, maakt de bijzin interessant om de positie van VzC's te bestuderen.

- (3) a [dat]_{Pool1} je zoiets wel **tegen Kaatje** [kunt zeggen]_{Pool2}.
b [dat]_{Pool1} je zoiets wel [kunt zeggen]_{Pool2} **tegen Kaatje**.

Diachrone studies hebben aangetoond dat de positie van VzC's al sinds het (vroeg) Middelnederlands varieert (o.a. De Meersman 1980, 1988; De Schutter 1988, 1995, 2003; voor een uitgebreide bespreking, zie §2.2.1). Toch kreeg de variabele positie van VzC's weinig aandacht in de algemene taalkundige literatuur, noch in de taalgebruiksgebaseerde (of empirische) onderzoekstraditie. De meeste studies naar dit type van grammaticale variatie spelen zich af in de generatieve onderzoekstraditie. In opeenvolgende voorstellen wil de generatieve taalkundige de onderliggende syntactische basisstructuur van het Nederlands (SOV of SVO) vastleggen en wil men aan de hand van transformaties en principes de afgeleide positie verklaren. De klemtoon in de formele onderzoekstraditie ligt dus steeds op de mentale representatie en het doel is om de taalfeiten zo adequaat en tegelijkertijd zo economisch mogelijk te genereren. Deze aanpak verschilt van de aanpak in onderhavige studie. In deze studie leggen we de klemtoon op de variabelen die de distributie tussen de twee woordvolgordevarianten in

³ Een alternatief is het subject als eerste pool in plaats van het voegwoord. Voor een bespreking, zie Zwart (2011: 33-37).

concreet taalgebruik kunnen verklaren. Het doel van deze synchrone empirische studie is dus om inzicht te geven in de precieze gebruiksvoorwaarden van de twee mogelijke posities van VzC's, die in principe vrij naast elkaar gebruikt kunnen worden.

Hoewel dit type van syntactische variatie geen uitgebreide onderzoekstraditie kent, zijn er toch enkele verklaringen geponeerd die de achterveldpositie van VzC's (ook wel achteropplaatsing of extrapositie⁴) kunnen motiveren. De *Algemene Nederlandse Spraakkunst* (afgekort ANS) verwijst daarvoor naar twee principes: het pragmatische of thematische principe van de informatiegeleding en het complexiteitsprincipe (Haeseryn et al. 1997: 1235-1246; deze principes worden uitgebreid beschreven in §2.1.1). Het eerste principe zegt simpelweg dat de informatiestatus of -waarde de keuze tussen middenveldpositie en achterveldpositie beïnvloedt. Het tweede principe zegt dat de afstand tussen de twee polen (i.e. het middenveld) niet te groot mag worden, aangezien taalgebruikers anders te lang moeten wachten op het semantisch belangrijke hoofdwerkwoord, waardoor de zin moeilijker te analyseren en te verwerken is. De invloed van deze principes op de variabele positie van VzC's werd ook empirisch onderzocht, meer bepaald door Jansen (1978, 1979) en Braecke (1990). Zij bestudeerden de positie van VzC's in het mondelinge taalgebruik van respondenten uit respectievelijk Leiden en Opwijk. Jansen (1978) toont o.a. aan dat complexere VzC's significant vaker in achterveldpositie staan, in overeenstemming met het Language-Preferred Order of Constituents (Dik 1978). Jansen (2003) verklaart dit ook psycholinguïstisch: lange constituenten staan in het achterveld om het middenveld (of de tangconstructie) te verlichten (cf. Hawkins Principle of Early Immediate Constituents 1994 en Gibsons Dependency Locality Theory 1998, 2000). De variabele *complexiteit* van de VzC speelt ook volgens Braecke (1990) een rol, maar deze rol is ondergeschikt aan een pragmatische variabele⁵ *focus*, overeenkomstig het Code-Quantity Principle van Givón.

Op basis van het eerdere synchroon empirisch onderzoek van voornamelijk Jansen (1978, 1979, 2003) en Braecke (1990) dringen zich twee cruciale bedenkingen op (zie ook §2.2.4 voor een uitgebreid overzicht en andere (methodologische) bedenkingen). Ten eerste zijn beide onderzoeken monofactorieel, d.w.z. dat het effect van een enkele variabele op het plaatsingsgedrag van de VzC bestudeerd wordt, waardoor het onduidelijk is hoe de verschillende variabelen of principes zich tot elkaar verhouden in

⁴ De plaatsing van elementen achter de werkwoordelijke eindgroep wordt in de literatuur doorgaans achteropplaatsing (o.a. ANS 1997: 1365) of extrapositie (o.a. De Vries 2011) genoemd. De term extrapositie verwijst oorspronkelijk naar een transformationele verplaatsingsregel uit de generatieve grammatica maar wordt tegenwoordig ook in meer traditionele volgordebeschrijvingen gebruikt als aanduiding van het verschijnsel zonder meer (cf. Vandeweghe 2000: 255-257). In deze bijdrage zullen beide termen naast elkaar gebruikt worden zonder hier theoretische consequenties aan te verbinden.

⁵ Ook Haeseryn (1998) valt de bedenking van Braecke (1990) bij, met name dat *lengte/complexiteit* ondergeschikt is aan *discoursstatus*.

hun effect op deze woordvolgordevariatie en of de principes in staat zijn de volledige variatie in plaatsingsgedrag te bepalen. Ook de ANS laat in het midden of en hoe de variabelen elkaar al dan niet kunnen beïnvloeden. Het is m.a.w. nog een open vraag of beide principes in dezelfde mate de positie van VzC's bepalen, zoals Jansen (1978) beweert, of één principe domineert, wat volgens Braecke (1990) het geval is. Ten tweede beperken de empirische studies naar de variabele positie van VzC's zich tot een gesproken corpus. We kunnen ons dus afvragen of dezelfde mechanismen op eenzelfde manier werkzaam zijn in een geschreven corpus. Anders geformuleerd: het is goed mogelijk dat andere, nieuwe variabelen, die misschien vanuit een gesproken perspectief niet onderzocht werden, een rol spelen in geschreven corpusmateriaal. Voorliggende studie stelt zich dan ook als doel om op een kwantitatief-corpusgebaseerde manier na te gaan of, en, zo ja, in welke mate de twee principes een rol spelen bij de bepaling van de woordvolgorde van VzC's in een geschreven corpus en wat de relatie is tussen deze twee principes. Daartoe zullen we:

- gebruikmaken van de journalistieke teksten binnen het Dutch Parallel Corpus (Macken et al. 2011);
- de twee principes op verschillende manieren operationaliseren en samenbrengen in één multivariate studie;
- het effect van de variabelen verifiëren, nagaan of ze interageren en eventuele verschillen met eerder onderzoek aanstippen en interpreteren.

Door geschreven journalistieke teksten te bestuderen, bieden we een tegengewicht aan de voorgaande empirische onderzoeken en kunnen we nagaan of de (*syntactische*) *complexiteit* en de *discoursstatus* wel werkzaam zijn in een geschreven, sterk geëditeerde en homogene⁶ tekstsoort zoals journalistieke teksten. Het spreekt voor zich dat taalgebruikers in een spontane gesproken situatie, waarbij de uitingen opeenvolgend geformuleerd worden (Levelt 1989; Wheeldon & Lahiri 1997; Ferreira & Swets 2002; Allum & Wheeldon 2007), die volgorde kiezen die het minst geheugenbelasting vraagt (Hawkins 1994; Arnold et al. 2000; Gibson 2000; Jansen 2003). We zouden kunnen verwachten dat sprekers eerst de middenveldpositie voor VzC's kiezen, tenzij het middenveld zelf te zwaar is (of wordt door de VzC). In dat geval kan de spreker ervoor opteren om een VzC naar het achterveld te verplaatsen (cf. ANS 1997). Vanuit een geschreven uitgangspunt kunnen we de voorgaande redenering in twijfel trekken. Ten eerste is het idee om een VzC uit te stellen tot het einde van een zin begrijpelijk in gesproken taal, maar daarom niet in geschreven en sterk geëditeerde taal. Schrijvers hebben namelijk de tijd om tijdens

⁶ Met *homogeen* verwijzen we naar het specifieke doel dat journalistieke teksten voor ogen hebben, namelijk de lezer informeren en opiniëren over actuele gebeurtenissen.

het schrijven of nadat ze een zin of tekst geschreven hebben, een VzC van plaats te veranderen en lezers hebben de tijd om even terug te springen in de zin als ze iets niet begrepen hebben. Dat zou ertoe kunnen leiden dat schrijvers minder rekening houden met het verlichten van de tangconstructie. Ten tweede impliceert de (optionele) beweging van het middenveld naar het achterveld in een gesproken situatie dat er enkel sprake kan zijn van een eenrichtingsbeweging: een VzC kan van het middenveld naar het achterveld verplaatst worden, maar niet omgekeerd. Dat is opnieuw heel plausibel vanuit een gesproken standpunt, maar daarom niet voor een medium waarbij de taalgebruiker tijd heeft om zijn output te herwerken. Andere variabelen, zoals de *lengte van het achterveld*, zouden ook de keuze tussen middenveld- en achterveldpositie kunnen beïnvloeden. In dat geval is het achterveld geen tweederangspositie of ‘slechts’ een toevluchtsoord voor een te zwaar beladen middenveld, maar een autonome positie die eveneens de plaatsingsvoorkeur van VzC’s bepaalt. Onze verwachting is dat er in geschreven journalistieke teksten bidirectioneel gewerkt wordt en dat er een wisselwerking bestaat tussen de *complexiteit van het middenveld*, de VzC en het *achterveld*. Om dit idee te staven, zullen we de variabelen *syntactische complexiteit van het middenveld*, de VzC en het *achterveld* aan een multivariate analyse toevoegen en gaan we na of deze variabelen interageren.

Deze studie leest zich als een verdere aanvulling, uitbouw en synthese van de bestaande literatuur. Door statistische technieken (zoals een logistische-regressieanalyse) te gebruiken, kunnen we eveneens bepalen welke variabele(n) het best een voorkeurspositie (middenveld of achterveld) verklaren en voorspellen, wat ook de huidige kennis over deze woordvolgordevariatie binnen de neerlandistiek zal verrijken. Deze studie zal bijgevolg een accurater en vollediger inzicht geven in de onderliggende mechanismen die de plaatsingsvoorkeur van VzC’s bepalen in geschreven taal. Bovendien vindt dit onderzoek zo ook aansluiting bij andere onderzoeken naar syntactische variatiefenomenen in het Nederlands, die wel een statistisch onderbouwde onderzoekstraditie kennen, zoals dat het geval is voor scrambling (van Bergen & de Swart 2010), de datiefalternantie (Colleman 2010), de groene en rode volgorde (De Sutter 2007; Coussé 2008) en de positie van presentatief *er* (Grondelaers & Speelman 2007; Grondelaers, Speelman & Geeraerts 2008; Grondelaers et al. 2009).

Om de repliceerbaarheid en continuïteit in toekomstig onderzoek te garanderen, moeten de onderzochte zinnen, VzC’s en de onderzochte variabelen aan strikte selectiecriteria beantwoorden (§3.1.2). Zo focussen we bijvoorbeeld op bijzinnen, om de evidente reden dat de tweede pool in bijzinnen steeds zichtbaar is en er bijgevolg een duidelijk onderscheid is tussen een VzC in het middenveld en een in het achterveld. Omdat het type bijzin een rol kan spelen, beperken we dit onderzoek tot voegwoordzinnen ingeleid door het semantisch lege, grammaticaal onderschikkende voegwoord *dat*. De zinnen kunnen probleemloos met elkaar vergeleken worden en tegelijkertijd beperken we de dataset enigszins. Voorts verdelen we de onderzochte

variabelen in vier groepen: 1. *syntactische complexiteit van de VzC, het middenveld en achterveld*, 2. *Discoursstatus van de VzC*, 3. *grammaticale en semantische functies van de VzC* en 4. *grammaticale en semantische classificatie van het (hoofd)werkwoord*. De variabelen die deel uitmaken van de eerste drie groepen, die ook in mindere of meerdere mate in het empirische onderzoek van Jansen (1978, 1979) en Braecke (1990) aan bod komen, worden in deze studie op verschillende manieren geoperationaliseerd. De laatste groep *grammaticale en semantische classificatie van het (hoofd)werkwoord* is nieuw, maar komt wel meermaals aan bod in andere onderzoeken naar syntactische variatiefenomenen in het Nederlands en het Engels (o.a., Stallings, MacDonald & O'Seaghdha 1998; Gries 2004; Bresnan et al. 2005; Bresnan & Ford 2010, Grondelaers et al. 2009, Hoeksema 2014). De operationalisatie van het werkwoord kan ook implicaties hebben voor de neerlandistiek en meer bepaald voor de verklaringen in de ANS. Indien de werkwoorden (of subklassen van werkwoorden) een individuele voorkeur hebben voor middenveldpositie of achterveldpositie, dan toont deze studie aan dat de woordvolgorde van VzC's in geschreven taal zich niet beperkt tot een verklaring in termen van een thematisch principe en/of complexiteitsprincipe.

Nu de doelstelling en opbouw van dit onderzoek duidelijk zijn, vatten we kort de verdere opbouw van dit werk samen. In het volgende hoofdstuk geven we een overzicht van gerelateerd onderzoek naar de positie van VzC's. Eerst bespreken we de algemene literatuur en meer bepaald de verklaringen en adviezen volgens de grammatica's en stijlguides. Ten tweede komt de wetenschappelijke literatuur aan bod. Binnen deze paragraaf presenteren we diachroon-historische studies naar de woordvolgordevariatie van VzC's, generatieve studies naar extrapositie, (functioneel) grammaticale studies en synchrone empirische studies, zoals dat van Jansen (1978, 1979) en Braecke (1990). Hoofdstuk 3 bespreekt het corpus en de manier waarop de data en de variabelen in deze studie geanalyseerd zijn. In hoofdstuk 4 presenteren we eerst een univariaat overzicht van de verschillende variabelen en vergelijken we de verschillende operationalisaties voor *syntactische complexiteit* en *discoursstatus*. Vervolgens voegen we de variabelen aan een multifactoriële statistische analyse toe. We stellen de variabelen die (g)een significant effect hebben voor. De multifactoriële analyse vertelt ons ook hoeveel variatie we op basis van de onderzochte variabelen kunnen verklaren en voorspellen. In het vijfde hoofdstuk koppelen we de resultaten aan bestaande ideeën en theorieën over de positie van VzC's en de functie van syntactische variabiliteit in het algemeen. In dit hoofdstuk zullen we dus de verklarende principes volgens de ANS, met name het thematische principe (*discoursstatus*) en het complexiteitsprincipe (*syntactische complexiteit*), en de andere onderzochte variabelen kunnen bevestigen, weerleggen en/of nuanceren. Daarnaast verklaren we de waargenomen trends op een functioneel-cognitieve manier en focussen we op de sterktes en tekortkomingen van dit onderzoek. Tot slot geven we ook aan op welk vlak vervolgonderzoek nuttig zou zijn en waarmee toekomstig onderzoek naar de positie van VzC's rekening moet houden.

Hoofdstuk 2

Literatuurstudie

In dit hoofdstuk geven we een overzicht van de bestaande algemene en taalwetenschappelijke literatuur over de variabele positie van VzC's. We presenteren eerst de verklaringen en adviezen van grammatica's en stijlgidsen, om te weten te komen hoe in het algemeen tegen de variatie aangekeken wordt. Vervolgens bespreken we onderzoek naar de positie van VzC's in de taalwetenschappelijke literatuur. Vanaf de jaren '70 ontstaat er in verschillende taalwetenschappelijke stromingen interesse voor de variabele positie van VzC's. Dat betekent geenszins dat er voor die tijd geen aandacht werd besteed aan de positie in het midden- en achterveld, wel bleef de aandacht meestal beperkt tot een vermelding van de mogelijkheid tot alternantie in de algemene literatuur (met name in grammatica's), hier en daar vergezeld door een voorkeur voor een van de woordvolgordes.¹

Hoewel deze dissertatie naar de positie van VzC's synchroon en empirisch te werk gaat, willen we in deze paragraaf een zo volledig mogelijk overzicht geven van alle taalwetenschappelijke studies die zijn verschenen over dit type van syntactische variatie. We verdelen deze paragraaf onder naargelang het type onderzoek: 1. diachroon-historische studies, 2. generatieve studies, 3. (functioneel) grammaticale studies en 4. synchroon empirische (of taalgebruiksgebaseerde) studies. Het spreekt voor zich dat onderhavige studie deel uitmaakt van het laatste type onderzoek.

¹ De reden waarom de positie van VzC's in het verleden zo weinig aan bod komt, is volgens Den Hertog omdat de invloeden die deze schikking kunnen bepalen heel subjectief zijn (Den Hertog 1973: 136). Ook De Schutter & van Hauwermeiren (1983) en Coussé (2009) wijzen vooral op de moeilijkheden van zo'n onderzoek. Het bepalen van wat de beste volgorde mag zijn, is volgens De Schutter & van Hauwermeiren een kwestie van heel fijne intuïtie en dat maakt het bijna onmogelijk om te formuleren waarom een bepaalde volgorde nu beter of minder goed is dan een andere. "In alle gevallen zal de context en de bedoeling van de spreker/schrijver een bepalende rol spelen" (De Schutter & van Hauwermeiren 1983: 194).

2.1 Algemene literatuur

In dit hoofdstuk gaan we na of, en in welke mate, de algemene literatuur de plaatsingsvariatie van VzC's bespreekt. We maken daarbij een onderscheid tussen geraadpleegde grammatica's (2.1.1) en stijlguiden (2.1.2). De onderzochte literatuur hebben we in een aparte literatuurlijst opgenomen (Appendix A). De grammatica's die de hier bestudeerde variatie vermelden, hebben we in Appendix A gemarkeerd met een asterisk.

2.1.1 Grammatica's

De grammatica die het uitvoerigst ingaat op de variabele positie van VzC's is de *Algemene Nederlandse Spraakkunst* (afgekort ANS). Volgens de ANS is het helemaal niet ongevoel om elementen na de werkwoordelijke eindgroep (of buiten de zogenaamde tang, in extrapositie of in achterveldpositie) te plaatsen. Er wordt wel een onderscheid gemaakt tussen "1. die gevallen waarbij achteropplaatsing gegrammaticaliseerd is, dat wil zeggen tot regel of sterke tendens geworden is, en 2. die gevallen die in principe zowel achteropplaatsing als niet-achteropplaatsing toelaten, maar waarbij in de praktijk diverse factoren een rol spelen bij de keuze" (ANS 1997: 1365). Een voorbeeld van de eerste categorie zijn afhankelijke zinnen. Afhankelijke zinnen staan bij voorkeur niet in het middenveld en in heel wat gevallen is het middenveld zelfs uitgesloten (zoals in zin 4a; ANS 1997: 1366). De alternatieve posities zijn het voorveld (zin 4b) en het achterveld (zin 4c), naast de aan- en uitloop. Omdat afhankelijke zinnen een sterke tendens vertonen om in het achterveld te staan (soms is het de enige mogelijke positie), is het achterveld volgens de ANS de basispositie voor dit soort zinnen (ANS 1997: 1367).

- (4) a Hij [komt] omdat hij erg boos is niet ||. <<uitgesloten>>
b Omdat hij erg boos is, [komt] hij niet ||.
c Hij [komt] niet ||, omdat hij erg boos is.

Een VzC is niet gebonden aan een bepaalde positie en is daarom het voorbeeld bij uitstek voor de tweede categorie, de optionele extrapositie. De ANS gebruikt twee principes om de positie van een VzC in het achterveld te verklaren: 1. het pragmatische of thematische principe van de informatiegeleding en 2. het complexiteitsprincipe (Haeseryn et al. 1997: 1235-1246). De ANS suggereert ook dat er een beweging is vanuit het middenveld naar het achterveld, of anders gezegd, dat het middenveld de zogenaamde basispositie is voor de VzC en dat het achterveld enkel in bepaalde situaties gebruikt zal worden. Net die 'situatie' wil de ANS met behulp van de variabelen

informatiewaarde (wij gebruiken in onderhavige studie de term *discoursstatus*) en de *complexiteit* verklaren.

De *discoursstatus* kan volgens de ANS op twee manieren een rol spelen. Zowel zinsdelen (en zinsdeelstukken) met grote informatieve waarde (zie zin 5a) als elementen met geringe informatieve waarde (zin 6) kunnen net door deze informatiestatus achteraan geplaatst worden.

- (5) a Hij [moet]_{Pool1} naar Zwitserland [gaan]_{Pool2} **om gezondheidsredenen**.
b Hij [moet]_{Pool1} **om gezondheidsredenen** naar Zwitserland [gaan]_{Pool2}.
- (6) Er [schijnt]_{Pool1} weer een spectaculair ongeluk [gebeurd te zijn]_{Pool2} **op de E17**.

Een VzC met een relatief grote informatieve waarde zal volgens het links-rechts-principe zover mogelijk rechts geplaatst worden: ofwel rechts in het middenveld, ofwel rechts in de zin, in het achterveld. Het links-rechts-principe stelt namelijk dat er een opgaande lijn is in een zin: “wat informatief belangrijker is, heeft de neiging meer achterin in de zin te komen” (ANS 1997: 1238). Als de VzC *om gezondheidsredenen* informatief belangrijker is, kan deze enkel achteraan geplaatst worden, zoals in zin 5a, aangezien het richtingscomplement *naar Zwitserland* inherent is en bijgevolg altijd vlak voor de tweede pool staat. Het informatieve verschil tussen 5a en b is dat in zin 5a de reden van de reis het informatieve zwaartepunt is, terwijl in zin 5b de kern van de informatie de plaats is waar de hij-persoon naartoe gaat. Althans bij een neutrale accentuering, want “men kan het informatief belangrijke element meer naar links plaatsen (zoals volgorde 5b) en het door een accent als informatief belangrijk karakteriseren” (ANS 1997: 1371). Het is met andere woorden niet zozeer de woordvolgorde maar het accent dat bepaalt wat informatief in het middelpunt staat.

VzC's kunnen ook om hun kleine informatieve waarde geëxtraponeerd worden (zoals in zin 6). Volgens de ANS gaat het meestal om plaats- of tijdsaanduidingen (ANS 1997: 1373) die als een bijkomstigheid gepresenteerd worden. Dat betekent dat een ander element in het middenstuk informatief belangrijker is of dat het werkwoord in de tweede pool informatief in het middelpunt staat. Zo kan zin 7a volgens de ANS twee interpretaties krijgen: “met een accent op *Parijs* is de kern van de mededeling ‘naar Parijs vliegen’. Ligt het accent op *vliegen*, dan is de werking zelf de kern van de informatie, bijvoorbeeld om een contrast aan te geven met een andere manier van reizen, zoals ‘met de trein gaan’. Alleen in dit tweede geval is de volgorde van 7b mogelijk” (ANS 1997: 1373).

- (7) a We [willen]_{Pool1} morgen **naar Parijs** [vliegen]_{Pool2}.
b We [willen]_{Pool1} morgen [vliegen]_{Pool2} **naar Parijs**.

De plaats van een constituent in de zin en het accent dat een taalgebruiker aan een bepaalde constituent geeft, kunnen dus aangeven of de informatie belangrijk is. Toch doet het feit dat zowel belangrijke als onbelangrijke informatie op dezelfde plaats kunnen staan, namelijk in het achterveld, vermoeden dat het accent de doorslaggevende factor is

die het laatste woord heeft. Het gesuggereerde verband tussen de informativiteit en de positie in de zin is m.a.w. niet altijd even rechtlijnig.

Het idee dat het accent bepaalt hoe de informatie begrepen moet worden, zou ook kunnen suggereren dat de ANS in de eerste plaats aan gesproken taal denkt. Een prosodisch accent is namelijk in eerste instantie een component van het spreekgedrag. Het is hoorbaar in een gesproken context en nauwelijks zichtbaar in een gedrukte tekst, want klemtonen en accenten schrijven we niet.² Dat roept de vraag op of het thematische principe, en dus de *discoursstatus* van VzC's, de positionering in de geschreven taal al dan niet beïnvloedt.

Het complexiteitsprincipe zegt dat de linguïstische afstand tussen de twee polen (i.e. de lengte en/of de complexiteit van het middenveld) niet te groot mag worden, aangezien taalgebruikers anders te lang moeten wachten op het semantisch belangrijke hoofdwerkwoord, waardoor de zin moeilijker te analyseren en te verwerken is. Om de afstand tussen de beide polen kleiner te maken ("om de 'tang' te verlichten" (ANS 1997: 1374)) wordt er een zinselement geëxtraponeerd. In deze verklaring ziet de ANS ook een stilistische verklaring: "De complexe bouw en/of de lengte van een element kan de twee polen van de zin zo ver uit elkaar plaatsen, dat de zin soms minder elegant wordt of aan verstaanbaarheid kan inboeten" (ANS 1997: 1366). Vooral in de gesproken taal zal men volgens de ANS sneller geneigd zijn om de VzC's achter de tweede pool te plaatsen. De spreker zal bijgevolg 8a boven 8b verkiezen (voorbeelden ANS 1997: 1374):

- (8) a Hij [had]_{Pool1} gelukkig uiteindelijk toch nog met veel moeite een flatje op de kop [kunnen tikken]_{Pool2} *in Valladolid*.
b Hij [had]_{Pool1} gelukkig uiteindelijk toch nog met veel moeite *in Valladolid* een flatje op de kop [kunnen tikken]_{Pool2}.

Zin 8a illustreert ook dat korte VzC's naar het achterveld verplaatst kunnen worden, als het middenveld door andere elementen te lang wordt. Naast de uitgebreidheid van het middenveld kan ook de uitgebreidheid van de VzC zelf het middenstuk lang maken (zie voorbeeld 9; ANS 1997: 1374):

- (9) a Ze [zou]_{Pool1} veel liever *in een van die leuke kleine dorpsschooltjes* les [geven]_{Pool2}.
b Ze [zou]_{Pool1} veel liever les [geven]_{Pool2} *in een van die leuke kleine dorpsschooltjes*.

² Dat wil echter niet zeggen dat een prosodisch accent de woordvolgorde in geschreven taal niet beïnvloedt. De Sutter et al. (2007) tonen bijvoorbeeld op basis van een hele reeks corpusanalyses aan dat de volgorde van het deelwoord en het hulpwerkwoord in geschreven werkwoordelijke eindgroepen beïnvloed wordt door het zgn. Principe van Ritmische Alternantie (PRA), i.e. de tendens van taalgebruikers om een accentenbotsing en een te lang interval met onbeklemtoonde lettergrepen te vermijden (2007: 19).

De ANS merkt verder nog op dat met het concept *uitgebreidheid* niet zozeer de mathematische lengte bedoeld wordt (i.e. het aantal woorden), maar wel de complexe bouw van een zinsdeel, bijvoorbeeld wanneer na het voorzetsel een naamwoordelijke constituent volgt, die zelf een betrekkelijke bijzin als bijvoeglijke nabepaling heeft (ANS 1997: 1366).

Kort samengevat, kunnen we concluderen dat de verklaring in termen van *complexiteit* eenduidiger is dan die in termen van *discoursstatus*. Toch valt het op dat de ANS ook bij het complexiteitsprincipe veeleer vanuit een gesproken context vertrekt. De taalgebruiker heeft volgens de ANS voornamelijk in een gesproken situatie de neiging om de verplaatsingsregel (of beter: achteropplaatsregel) toe te passen. Daarnaast vermeldt de ANS dat de *informatieve waarde* en de *complexiteit* elkaar al dan niet kunnen versterken, maar blijft de manier waarop deze twee principes interageren niet duidelijk. In dit onderzoek proberen we de precieze invloed van beide principes en de interactie ertussen in een empirisch-kwantitatieve studie te meten.

Naast de ANS werden nog 45 andere grammatica's bestudeerd. 23 daarvan gaan niet in op de woordvolgordevariatie, 22 wel. Van deze 22 grammatica's zijn er 4 die de variabele positie louter vermelden maar niet verklaren: Paardekooper (1963); De Schutter (1974); De Schutter & van Hauwermeiren (1983); Luif (1998). De Schutter (1974: 53), bijvoorbeeld, vermeldt zin 10 en voegt hieraan toe dat de VzC in *Antwerpen* geen zelfstandige betekenis heeft en dat het enkel "een marginale (voor het geheel als zodanig niet relevante bepaling) toevoegt" (De Schutter 1974: 55). De Schutter verklaart echter niet waarom dat zo is en of dit voor elke zin met een VzC in extrapositie geldt.

(10) Hij [heeft]_{pool1} op kamers [gewoond]_{pool2} **in Antwerpen**.

Van de overige 18 grammatica's verwijzen er 9 naar de accentuering van het zinsdeel(stuk) in extrapositie. Het idee dat een VzC in extrapositie meer *accent*³ krijgt, vinden we terug bij Hasselbach (1905); Overdiep (1928, 1949); Fraussen (1939); De Groot (1949); Vindevogel (1958); De Vooys & Schönfeld (1967); Van den Berg (1967); Smedts & Van Belle (2011). Voor Van den Berg is een afwijkende zinsbouw een manier om de woorden "een grotere kracht van uitdrukking te geven" (Van den Berg 1967: 45). Met 'afwijkend' bedoelt Van den Berg een VzC in het achterveld (zin 11a uit Van den Berg 1967: 45) tegenover een neutrale zinsconstructie (zin 11b) met eenzelfde VzC in het middenveld⁴. Voor Vindevogel is de laatste zinsplaats de meest expressieve plaats in de

³ In de geraadpleegde grammatica's worden ook nog de begrippen *klemtoon*, *aandacht* of *nadruk* gebruikt. We beschouwen ze in dit overzicht als synoniemen voor *accent*.

⁴ In Van den Berg (eerste druk 1952) wordt een VzC in achterveldpositie nog expliciet afgekeurd. Enkel bij expressief en herhalend gebruik kan de VzC in het achterveld staan.

zin, waarbij ‘expressief’ gelezen moet worden als “een ongewone plaats” om op “een woord(groep) speciaal de aandacht te vestigen” (Vindevogel 1958: 171). Uit het idee dat het achterveld een afwijkende of ongewone positie is, kunnen we ook afleiden dat het middenveld volgens Van den Berg⁵ en Vindevogel de meest neutrale basispositie is.

- (11) a Wat je voor mij doet, [moet]_{Pool1} helemaal [zijn]_{Pool2} **uit vríje wil**. (afwijkend)
 b Wat je voor mij doet, [moet]_{Pool1} helemaal **uit vrije wil** [zijn]_{Pool2}. (neutraal)

Vindevogel merkt ook op dat “de andere zinsdelen, die niet verplaatst werden, meer reliëf krijgen” (Vindevogel 1958: 171). In zin 12b krijgt zowel *mijn boek kwijt* als *sedert verleden week* meer reliëf⁶ in vergelijking met zin 12a (voorbeelden overgenomen uit Vindevogel 1958).

- (12) a Ik [ben]_{Pool1} **sedert verleden week** mijn boek kwijt [_{Pool2}]. (neutraal)
 b Ik [ben]_{Pool1} mijn boek kwijt [_{Pool2}] **sedert verleden week**. (expressief)

Vindevogel en Hasselbach maken net zoals de ANS de kanttekening dat de toon (Hasselbach 1905: 34-35) of de accentuering (Vindevogel 1958: 171) elke bepaling meer of minder kan doen uitkomen, ongeacht de positie in de zin.

Het tegenovergestelde idee, namelijk dat een VzC als een bijkomstigheid in het achterveld staat, om zo meer aandacht te geven aan een ander element in het middenveld of aan het werkwoord in het bijzonder (zie ook ANS 1997: 1373), vinden we terug bij Rijpma en Schuringa (1978) en Paardekooper (1986). Zij wijzen op de rol van het accent en hoe dat kan verklaren waarom sommige VzC’s uitgesloten zijn van een positie in het achterveld. In zin 13 (uit Paardekooper 1986: 28) ligt het accent op het werkwoord *komen* en kan de VzC *om acht uur* zowel pre- als postverbaal staan. Postverbaal zal het werkwoord een contrastaccent krijgen, bijvoorbeeld: *dat we kómen om acht uur* en niet *vertrékken om acht uur*. In zin 14 krijgt *om acht uur* het zinsaccent en kan het volgens Paardekooper enkel preverbaal staan. Rijpma en Schuringa spreken over “eindstukzinsdelen” die geen afzonderlijk accent verdragen (1978: 237).

- (13) a Als we **om acht uur** kómen.
 b Als we kómen **om acht uur**.
 (14) a Als we **om ácht uur** komen.
 b *Als we komen om ácht uur.

Van de 18 grammatica’s verwijzen er 7 naar een verklaring in termen van *complexiteit*: Overdiep (1928, 1949); De Vooy & Schönfeld (1967); Den Hertog (1973); Van den Toorn

⁵ Hoewel Van Haeringen (1947) het advies geeft om tangconstructies te vermijden, neemt Van den Berg (1949) een geheel andere positie in: ‘Het is regel in de Nederlandse zin dat datgene wat krachtens de betekenis het nauwst verband houdt, zo ver mogelijk uiteen geplaatst wordt.’ (uit Renkema 1989: 444 – 445).

⁶ Vindevogel specificeert het begrip *reliëf* niet, vandaar de aanhalingstekens.

(1981); Vandeweghe (2000); Smedts & Van Belle (2011). De Vooy's & Schönfeld hebben het daarbij voornamelijk over de complexiteit van het middenveld – extrapositie krijgt de voorkeur als “door te vele of te lange bepalingen de afstand tussen subject en verbum wat te lang zou worden” (De Vooy's & Schönfeld 1967: 393) – terwijl Van den Toorn op de lengte van het zinsdeel zelf focust (1981: 62). Volgens Overdiep (1928 en 1949) is extrapositie dan weer zeer begrijpelijk, “wanneer op de voorzetselbepaling een bijzin volgt” (cf. *uitgebreidheid* ANS 1997: 1366; Den Hertog 1973: 140; zie voorbeeld 15), “of wanneer de voorzetselbepaling zelf door zijn omvang den zin te zeer zou overladen en er nadruk op wordt gelegd” (Overdiep 1928: 212; Overdiep 1949: 540). In dat laatste geval gaat het om een combinatie van *complexiteit* en *accentuering* (voorbeeld 16).

- (15) Ik [heb]_{Pool1} [gezegd]_{Pool2} **tot de vrouw, die mijn zorgen deelt.** (Multatuli)
 (16) Voor de open deur [stond]_{Pool1} een meid de stoep [te schrobben]_{Pool2} **met opgeschorten rok en mouwen omhooggeschort om haar rooden armen.** (Aug. de Wit)

Volgens Smedts & Van Belle is het trouwens ook beter voor de *leesbaarheid* om de afstand tussen de delen van de tang niet te groot te laten worden. “Een dergelijke ‘overspannen’ tang kan het best doorbroken worden door zinsdelen naar voren of naar achteren te verplaatsen of door een andere zinsconstructie te gebruiken” (Smedts & Van Belle 2011: 373). De verklaring in termen van *leesbaarheid* valt binnen het stilistische plaatje dat de ANS eveneens aan de complexiteit verbindt.

Behalve de *accentuering* en de *complexiteit* vinden we in vijf geraadpleegde grammatica's ook een verklaring op basis van de *tekstsoort*, met name spreektaal en literaire taal. Het eerste is het geval bij Overdiep (1928,1949); Van Es (1966); De Vooy's & Schönfeld (1967); Olbrechts (1966). “De open zinsvormen (i.e. een zinselement in extrapositie) zijn in de omgangstaal gebruikelijk bij aarzelend spreken, bij een aanvullende spreekwijze. Door de open zinsvorm rijgt de spreker zijn gedachten, stukje voor stukje aaneen” (Overdiep 1928: 207, Overdiep 1949: 524-525). Een verwijzing naar de literaire taal vinden we terug bij Overdiep (1928, 1949) en bij Van Es (1966; Van Es heeft het over de ‘gestileerde’ taalsoort). De literaire schrijver is volgens Overdiep meer vrij in de plaatsing van een zinselement aan het einde van de zin. “Dat kan door de individuele consequenties ten opzichte van de natuurlijke ‘vrije’ vormen van woordschikking en syntaxis, die den individueelen stilist eigen zijn, verklaard worden” (Overdiep 1949: 537). Deze schikking is voornamelijk aantrekkelijk “door den dalenden en rijzenden rhythmischen vorm” (zie zin 17) “en om de overzichtelijkheid van den open zinsvorm, waar eindeloos bepalingen kunnen worden aangehecht in de beschrijving” (zie zin 18; Overdiep 1928: 212).

- (17) En dan zullen er wézen ìn dit lánd, die hem gróeten met líéfde, en die de hánd zullen léggén òp zijn hóofd (Multatuli)

- (18) Zoo dommelde hij voort, droomend van zijn eigen Meer en van hoge ijsbergen, die langzaam, langzaam voortschoven over den dijk, over het kerkhof, door de kerk heen, tegen de school, door de school heen,... (Eigenhuis).

Andere grammaticawerken zijn in hun verklaring niet altijd even duidelijk. Van Bart, Kerstens & Sturm (1998), bijvoorbeeld, verwijzen naar de vorm van de VzC, de betekenis ervan en ook naar het werkwoord waar de VzC bij hoort (Van Bart, Kerstens & Sturm 1998: 178). Ze expliciteren echter niet welke ‘vorm’, ‘betekenis’ en op welke manier het ‘werkwoord’ een rol kunnen spelen. Tacx (1976) stelt dat er na de tweede pool (of de werkwoordelijke eindgroep) nog een bepaling kan volgen, maar enkel als de spreker daarmee een speciale bedoeling heeft (Tacx 1976: 177). Die ‘speciale bedoeling’ blijft opnieuw ongedefinieerd. Den Hertog (1973) geeft naast de complexiteit nog een tweede verklaring die zegt dat de taalgebruiker die woordvolgorde moet kiezen die aan de lezer direct duidelijk maakt op welk woord de nadruk moet vallen (Den Hertog 1973: 140). Hij expliciteert niet welke volgorde dat is of welke omstandigheden bepalen waar de nadruk precies ligt. Bovendien ziet Den Hertog “weinig bezwaar in het weglaten van bijzonderheden over dit onderwerp (i.e. de plaatsbepaling van voorwerpen en bepalingen in een zin, en dus ook in het midden- en achterveld), omdat een door goede lectuur ontwikkelde taalzin hier het best de weg wijst” (Den Hertog 1973: 140).

Naast grammatica’s die in eerste instantie op moedertaalsprekers gericht zijn, werden ook 4 grammaticale lexica in beschouwing genomen. Enkel Van Bart & Sturm (1987) geven een overzicht van welke constituenten in extrapositie kunnen staan. De vraag waarom iets geëxtraponeerd wordt, is volgens de auteurs “een vraag die (nog) niet exact te beantwoorden is en waar dus nog verder onderzoek voor nodig is” (van Bart & Sturm 1987: 209).

2.1.2 Stijlgidsen

In deze paragraaf geven we een overzicht van wat stijlgidsen over de middenveld- en achterveldplaatsing van VzC’s zeggen. Geen enkele van de 25 geraadpleegde stijlgidsen⁷ verklaart expliciet waarom een VzC in het middenveld of achterveld staat. Wel geven ze

⁷ De 25 geraadpleegde stijlgidsen zijn soms zeer verschillend wat de doelstelling en de inhoud ervan betreft. Zo zijn er zeker verschillen tussen de stijlboeken die kranten- en televisieredacties uitgegeven hebben, zoals het Stijlboek van Permentier & Van Den Eynden (1998), en Succesvol formuleren, een advies- en oefenboek voor professionele schrijvers van Verhoeven & Wekking (1991). Toch behandelen we de werken die allerlei taal- en/of stijladvies geven samen. Het spreekt vanzelf dat ook de hierboven besproken grammatica’s een zekere (normatieve) taaladviserende functie hebben (cf. ANS 1997: 5-6), maar ze verschillen van de hier behandelde taaladviserende werken doordat de grammatica’s een min of meer volledige en systematische behandeling van de Nederlandse grammatica geven en de focus in eerste instantie niet op taalzorgkwesaties en schrijfadvisen ligt.

het algemene advies om woorden die bij elkaar horen, zoals de twee polen in een hoofdzin of het onderwerp en het werkwoord in een bijzin, dicht bij elkaar te plaatsen. Kortom, ze bespreken hoe taalgebruikers met tangconstructies om moeten gaan en vertellen daardoor impliciet wat er moet gebeuren met andere zinslementen, zoals VzC's, in en rond de grijpers van een tangconstructie.

Het algemene advies om tangconstructies te vermijden kunnen we ook lezen als een oproep om zinslementen die in het achterveld kunnen staan (zoals niet-inherente VzC's) te verplaatsen. Dat betekent niet dat een VzC in het middenveld fout is of te allen tijde vermeden moet worden. Volgens Hendrickx (2003: 230) kan dit in extremis zelfs tot hilarische uitspraken leiden, zoals in voorbeeldzin 19b:

- (19) a dat de patiënten met hoofdpijn ter plaatste behandeld werden
b *dat de patiënten ter plaatse behandeld werden met hoofdpijn

Een tekst met veel tangconstructies is wel moeilijker te lezen en te verwerken, ook al is het volgens Drop & De Vries verleidelijk voor de schrijver. De schrijver kan ervoor kiezen om alles tussen de grijpers van de tang te plaatsen, om het “kwijt” te zijn of “niet meer vast te houden” (Drop & De Vries 1977: 222). Voor de lezer wordt het moeilijker: een tangconstructie dwingt nu eenmaal de lezer om woorden te onthouden en te herlezen en “dat leidt af van wat de schrijver wil zeggen” (Zuijdgeest 1995: 33; van der Horst 1996: 39). De cognitieve verwerkingskost is dus hoger bij een tangconstructie dan bij een onttangde constructie.

Toch zijn tangconstructies in heel wat gevallen onvermijdelijk. De twee polen in een hoofd- en bijzin staan nu eenmaal vaak uit elkaar. Bovendien kan het geheugen een zekere mate van afstand wel verwerken. Lezen gaat immers snel en de mogelijkheid om terug te kijken bestaat. Volgens Zuijdgeest zijn tangconstructies toegelaten “zolang de lezer de opbouw in één keer kan doorgronden” (Zuijdgeest 1995: 30). Ook de inhoudelijke voorspelbaarheid van het tweede deel speelt daarbij een rol. In zin 20 en 21 (uit resp. Verhoeven & Wekking 1991: 59 en Van der Horst 1996: 39) voelt de lezer inhoudelijk aankomen dat *het vandalisme* zal worden *aangepakt* of *bestreden* en dat de *vraagstukken* worden *uitgewerkt* of *opgelost*. In dergelijke gevallen kan de betreffende bijzin (zin 20) of de VzC (zin 21) toch in het middenveld blijven staan

- (20) Het wordt de hoogste tijd **dat het voetbalvandalisme**, dat er allengs voor begint te zorgen dat de echte liefhebbers van deze fantastische sport het stadion uitgedreven worden en dat burgemeesters meer en meer noopt wedstrijden te verbieden, **met harde hand wordt aangepakt**.
- (21) In dit rapport zullen **enkele belangrijke vraagstukken** op het gebied van de economische orde, en met name de vraag naar de gewenste omvang van de investeringen, **worden uitgewerkt**.

Een tangconstructie kan ook goede diensten bewijzen “wanneer een schrijver bepaalde informatie als eenheid wil presenteren of als het ware in het voorbijgaan wil noemen” (Renkema 1995: 53). Concreet betekent dit dat de taalgebruiker een VzC in het middenveld plaatst, net omdat die VzC informatief weinig bijdraagt. Renkema onderzocht daarvoor in drie experimenten de moeilijkheidsgraad van twee typen tangconstructies, de lidwoord-naamwoord-tang⁸ en de bijzin-bijzin-tang⁹, en de attentiewaarde van het middenstuk. Voor de moeilijkheidsgraad werd de leestijd van de tangconstructie en de onttangde constructie gemeten, werden er verificatievragen gesteld en werd ook de tijd gemeten die nodig was om die vragen te beantwoorden. Enkel bij de bijzin-bijzin-tang is er een significant verschil gevonden voor de leestijden van het middenstuk: het middenstuk vergt in de tangconstructie meer leestijd in vergelijking met de onttangde constructie. Ondanks de langere leestijd is er geen verschil in tekstbegrip (Renkema 1989). De attentiewaarde van het middenstuk is met behulp van tikfouten en inhoudelijke wijzigingen in het middenstuk getest. Volgens het onderzoek worden tikfouten minder vaak gedetecteerd in de tangconstructie. Renkema (1989) concludeert hieruit dat een schrijver voor een bijzin-bijzin-tangconstructie kan kiezen als de informatie in het middenstuk minder belangrijk is.

Verhoeven & Wekking (1991) wijzen op een verschil tussen geschreven en gesproken taal. Volgens hen is het vermijden van tangconstructies in gesproken taal “veel belangrijker dan in de geschreven taal. Luisteren gaat immers langzamer dan lezen en je kunt niet terugspoelen” (Verhoeven & Wekking 1991: 59). Zuiddgeest (1995) stelt nog een ander genre-effect vast: in kranten en tijdschriften staan volgens hem zelden tangconstructies, terwijl juridische teksten en teksten van de overheid er bol van staan.

2.1.3 Conclusie algemene literatuur

Op basis van bovenstaand overzicht kunnen we besluiten dat geen van de onderzochte algemene werken een van de syntactische varianten afkeurt, noch een VzC in het middenveld, noch een in het achterveld. Toch gaan grammatica's en stijlgidsen anders om met de variabele positie van VzC's.

De geraadpleegde grammatica's proberen voornamelijk te verklaren waarom een VzC naar het achterveld wordt verplaatst. De meeste grammatica's laten daarbij uitschijnen dat het middenveld de basispositie is. Hoewel de ANS het woord *basispositie* niet gebruikt, kunnen we uit de eenrichtingsbeweging van het middenveld naar het achterveld, die de

⁸ Ter illustratie: De langdurige echte minima verkeren op dit moment in een onmenselijke positie aldus *een door medewerkers van de Leidse universiteit geschreven rapport* (Renkema 1989: 447)

⁹ Ter illustratie: Het CITO deelde desgevraagd mee *dat, hoewel de score alleen een indicatie is voor een eventuele vervolgopleiding, de toets steeds meer de status krijgt van het belangrijkste basis-examen* (Renkema 1989: 447)

ANS suggereert, wel veronderstellen dat de VzC eerst in het middenveld staat en pas in bepaalde ‘situaties’ naar het achterveld verhuist. Twee verklaringen komen meermaals naar voren: ten eerste krijgt de VzC meer *accent* in het achterveld, waardoor het achterveld dé positie voor belangrijke informatie is (cf. het thematische principe); ten tweede kan een VzC in het achterveld een *complex middenveld* ontlasten (cf. het complexiteitsprincipe). Hoewel we volgens het thematische principe verwachten dat belangrijke, nieuwe informatie in het achterveld staat, maakt de ANS zelf de kanttekening dat ook onbelangrijke VzC’s net op die plaats kunnen komen te staan. Daardoor blijft het onduidelijk of belangrijke, dan wel onbelangrijke VzC’s de achterveldpositie innemen. Daarnaast vermeldt de ANS dat het thematische principe en complexiteitsprincipe elkaar kunnen beïnvloeden, maar in welke mate dat precies gebeurt en hoe we dat precies moeten begrijpen (welke variabele speelt in welke situatie een meer determinerende rol), blijft eveneens onduidelijk.

Stijlgidsen geven overwegend het advies om tangconstructies te vermijden en dus om een VzC uit het middenveld te houden en naar het achterveld te verplaatsen. Vanuit het perspectief van de spreker is het soms makkelijker om de VzC in het middenveld te plaatsen. In dat geval moet de taalgebruiker minder onthouden en is hij alle informatie direct kwijt. Vanuit het perspectief van de hoorder is het verwerkingsproces daardoor net zeer belastend. Stijlgidsen geven sprekers én schrijvers daarom het advies rekening te houden met dit cognitieve proces.

Tot slot merken we op hoe stijlgidsen de grijpers van de tang in een bijzin anders definiëren dan grammatica’s. Volgens grammatica’s zijn het bindwoord en de werkwoorden de grijpers van de tangconstructie. De meeste stijlgidsen vermelden daarentegen de tang *onderwerp-werkwoord*.¹⁰ We stellen m.a.w. vast hoe grammatica’s vooral vanuit een structuralistisch perspectief vertrekken, terwijl stijlgidsen vanuit een conceptueel-syntactisch perspectief vertrekken. Het subject staat misschien niet altijd structureel aan het begin van de bijzin, het is in tegenstelling tot het bindwoord wel conceptueel-syntactisch met het werkwoord verbonden. Nadat een taalgebruiker het subject heeft uitgesproken, moet hij het subject, of op zijn minst het getal van het subject, onthouden om op een correcte manier het vervoegde werkwoord met het subject te laten congrueren. Deze twee operationalisaties van de ‘tang’ zullen in dit onderzoek ook geoperationaliseerd (zie §3.2.1.3) en vergeleken worden (zie §4.1.2.5).

Met de algemeen theoretische bagage uit dit hoofdstuk in het achterhoofd gaan we in het volgende hoofdstuk in op de volgordevariatie van VzC’s in de (taal)wetenschappelijke literatuur.

¹⁰7 stijlgidsen vermelden de tang *subject-vervoegd werkwoord*, 4 werken de tang *subject-gezegde*. In dat laatste geval gaat het over de volledige werkwoordelijke eindgroep en niet alleen over het vervoegde werkwoord. Enkel Klein & Visscher (1992) vermeldt de tang *bindwoord-gezegde*.

2.2 Wetenschappelijke literatuur

In deze paragraaf maken we de sprong van de algemene literatuur naar de wetenschappelijke literatuur. In paragraaf 2.2.1 presenteren we eerst enkele historische onderzoeken. In de daaropvolgende paragrafen verschuift de focus van het historische naar het synchroon-hedendaagse onderzoek. De talrijke generatieve onderzoeken naar extrapositie worden in paragraaf 2.2.2 besproken. In paragraaf 2.2.3 bespreken we enkele onderzoeken binnen twee andere grammaticamodellen, meer bepaald de functionele (discourse) grammatica en de constructiegrammatica. Paragraaf 2.2.4 presenteert tot slot empirisch (taalgebruiksgebaseerde) onderzoek naar de variabele positie van VzC's in het Nederlands.

2.2.1 Historisch onderzoek

Historische studies (synchroon of diachroon) zijn van cruciaal belang om meer te weten te komen over de historische wortels van de moderne volgordevariatie. De hamvraag is of de variatie tussen middenveldplaatsing en achterveldplaatsing van VzC's ook in het verleden geldig was en of er een verandering ten opzichte van de huidige woordvolgordevariatie waargenomen kan worden. In deze paragraaf geven we eerst een algemeen overzicht van de constituenten die voor achterveldplaatsing in aanmerking kwamen (Coussé 2003; Van der Horst 2008). Vervolgens gaan we dieper in op historische studies die eveneens op zoek gaan naar de onderliggende mechanismen om de plaatsing van o.a. VzC's te verklaren (De Meersman 1980, 1988; De Schutter 1988, 1995, 2003; Coussé 2003, 2011; Hoeskema 2014). Met deze paragraaf willen we een antwoord geven op de volgende vragen:

- Kwamen VzC's in het verleden (en door de eeuwen heen) in aanmerking voor extrapositie?
- Welke variabelen hebben daarbij een rol gespeeld?

2.2.1.1 Diachroon overzicht van de zinsdelen in achterveldpositie

Voor een algemeen diachroon overzicht verwijzen we in de eerste plaats naar de *Geschiedenis van de Nederlandse Syntaxis* (Van der Horst 2008). Daaruit kunnen we concluderen dat de keuze om een zinselement in het achterveld te plaatsen al ver teruggaat in de tijd en dat de achterveldpositie vroeger veel vrijer was dan nu. De trend om zinselementen in het achterveld te plaatsen is met andere woorden afgenomen.

In het Middelnederlands konden volgens Van der Horst (2008) de volgende zinsdelen makkelijk na de werkwoordelijke eindgroep staan: voorzetselgroepen, bijzinnen, een (nominaal) subject, een direct object (soms een pronominaal direct object), een indirect object, een predicaatsnomen, een adverbium of adverbiale groep, een bepaling van gesteldheid, een genitief, delen van gesplitste werkwoorden en gesplitste voornaamwoordelijke bijwoorden (Van der Horst 2008: 553). Hoewel het in de 14^e eeuw dus heel gewoon was om allerlei zinsdelen na de werkwoordelijke groep te plaatsen, neemt die mogelijkheid in de loop van de 15^e eeuw af (Van der Horst 2008: 779). In de 16^e eeuw is extrapositie van andere elementen dan voorzetselgroepen, voegwoordgroepen, infinitiefgroepen en bijzinnen zelfs ongewoon (Van der Horst 2008: 1051) en in de 17^e eeuw zet deze tendens zich voort. De mogelijkheden om zinsdelen na de werkwoordelijke eindgroep te plaatsen, zijn in de 18^e eeuw volledig te vergelijken met het moderne Nederlands. Enkel in bijzondere gevallen, bij lange zinsdelen of bij een speciale intonatie, kunnen andere zinsdelen na de eindgroep staan (Van der Horst 2008: 1604).

Dat er in de 13^e en 14^e eeuw geen duidelijke voorkeurspositie was tussen midden- en achterveld bevestigt ook Coussé (2003). Met behulp van een corpus ambtelijke teksten uit de periode 1250-1800 geeft ze een overzicht van de achteropgeplaatste constituenten in verschillende bijzinnen en over verschillende eeuwen heen (2003: 157-158). Ze stelt vast dat vooral vanaf de 15^e eeuw het aandeel extrapositie gevoelig daalt: in de 13^e en 14^e eeuw kon gemiddeld meer dan de helft van de woorden achter de werkwoordelijke eindgroep geplaatst worden, in de loop van de 15^e eeuw daalt dit aantal tot een dertigtal procent. Bovendien konden in de 13^e eeuw tot zeven constituenten achter de werkwoordcluster staan, terwijl dat aantal terugloopt tot slechts één constituent in de 19^e eeuw (Coussé 2011).

Van der Horst (2008) wijst ook op aanzienlijke verschillen tussen auteurs. Bij Poirters en de Swaen, beiden zuidelijke auteurs, zijn er nog veel mogelijkheden na de eindgroep. Dat zou op een mogelijk Noord-Zuidverschil¹¹ kunnen wijzen, wat ook Burridges uitspraak ondersteunt: “Generally, it seems that Hollandish grammaticalizes its word order earlier than Brabantisch” (Burridge 1993: 61).

Het is duidelijk dat VzC's in het verleden en door de eeuwen heen in aanmerking kwamen voor extrapositie. Zelfs zinsdelen die momenteel van achteropplaatsing zijn uitgesloten, verschenen in extrapositie. Op basis van dit algemene overzicht kunnen we concluderen dat het achterveld in het Middelnederlands een autonome positie was waar meerdere zinsdelen in verschillende vormen konden staan.

¹¹ Van der Horst verwijst hierbij ook naar Vanacker en diens vaststelling dat in sommige West-Vlaamse dialecten tot op heden inherente voorzetselgroepen na de werkwoordelijke eindgroep kunnen staan (Vanacker 1977).

2.2.1.2 Variabelen die achterveldpositie verklaren

Enkele historische studies zijn op zoek gegaan naar de onderliggende mechanismen die de vrije achterveldplaatsing verklaren. Coussé (2003), bijvoorbeeld, vindt de verklaring in de veranderende plaatsingsmogelijkheden van de *zinsfocus* (cf. De Meersman 1988, De Schutter 1988 en Burridge 1993) en stelt een verschuiving vast van een zuiver discoursgemotiveerde invulling naar een minder vrije en meer gegrammaticaliseerde vorm van extrapositie. De positie achter de werkwoordelijke eindgroep was in het Middelnederlands de plaats bij uitstek voor de *zinsfocus*, d.w.z. constituenten die nieuwe belangwekkende informatie in het discours introduceren:

In short, exbraciated material is likely to be non-topical material; i.e. usually unknown information, that which [sic] cannot be understood from the context and which is not shared by the speaker and hearer. On the other hand, material which is not presupposed, and which is therefore the most assertive part of the utterance, is the most likely material to stand outside the brace. (Burridge 1993: 107)

Die doorgedreven discoursgemotiveerde invulling is in latere fases van het Nederlands onmogelijk geworden voor nominale constituenten. Volgens Coussé is in het moderne Nederlands de positie achter de werkwoordelijke eindgroep beperkt tot hoofdzakelijk complexe zindelen of zinsdeelstukken in de vorm van een bijzin of ingeleid door een voorzetsel.

De Meersman (1980, 1988) en De Schutter (1988, 1995, 2003) bestuderen enkel het Middelnederlands en gaan dieper in op de plaatsing van VzC's. De Meersman (1980) gebruikt daarvoor *de Grote Enquete*, een lange, vrij homogene ambtelijke tekst uit 1389 in het Zuidwestbrabants¹², om de plaatsing van VzC's tegenover de subject-werkwoord-tang¹³ in de betreffende en voegwoordelijke bijzin te bestuderen. De Schutter bestudeert het (vroeg) Middelnederlands aan de hand van 13^e-eeuwse teksten.

Beide auteurs vertrekken van een algemeen onderliggende SOV-volgorde¹⁴, hoewel de vrije plaatsing van VzC's deze volgorde in twijfel trekt. Volgens de Meersman (1980) is het Middelnederlands een SOV-taal, omdat de SxV-volgorde (of middenveldpositie) voor vrijwel alle zinsdelen (direct object, indirect object, bijwoordelijke bepaling en predicaatsnomen) domineert. Enkel niet-inherente VzC's als bijwoordelijke bepalingen kunnen makkelijk achterop staan, vooral als ze in een voegwoordelijke bijzin voorkomen. Toch blijft de ongemarkeerde positie voor deze VzC's het middenveld. In later onderzoek weerlegt de Meersman deze *gemarkeerdheid*: “noch de frequentie noch de

¹² Meer details zie de Meersman 1988: 103-104.

¹³ De Meersman (1980, 1988) houdt rekening met een subject-werkwoord-tang. Om deze keuze te verdedigen, verwijst de Meersman naar van Haeringen (1947, 1956) en naar Jansen (1978).

¹⁴ I.e. de volgorde Subject-Object-Verb

gebruiksvoorwaarden van die plaatsing (i.e. het achterveld) geven daar aanleiding toe (namelijk om de achterveldpositie als gemarkeerd te beschrijven)” (1988: 116). De Meersman besluit wel dat, ondanks de vrijheid om VzC’s in het midden- en achterveld te plaatsen, de positie voor het werkwoord de meest neutrale positie is voor alle zinsdelen en dus ook voor VzC’s (de Meersman 1988: 116), conform zijn uitgangspunt van een onderliggende SOV-volgorde. De Schutter (1988) laat in het midden of het Nederlands een “SOV-taal is met een recursief optredende optionele extrapositieregel”, dan wel “een gedeeltelijke SVO-taal in bijzinnen” (De Schutter 1988: 394). In latere onderzoeken gaat De Schutter (1995, 2003) wel telkens uit van een onderliggende SOV-volgorde voor het Middelnederlands.

Variabelen die volgens de Meersman (1980) VzC’s tot extrapositie kunnen stimuleren, zijn de *complexiteit* (de aanwezigheid van een betrekkelijke bijzin als nabepaling doet de SVx-volgorde of de achterveldpositie toenemen) en de aard van de *bijzin* (de SxV-volgorde is sterker in de relatiefzin dan in de voegwoordelijke bijzin). Andere variabelen die een rol zouden kunnen spelen, zijn de *twee- of meerledige werkwoordelijke groep* (twee- of meerledige werkwoordelijke groepen stimuleren middenveldpositie), *nevenschikking* (de kans op achterveldpositie neemt toe als een voegwoordelijke bijzin nevenschikkend verbonden is met een bijzin die al een zinselement in achterveldpositie heeft) en de *inherente plaats- en richtingsbepalingen* (inherente plaats- en richtingsbepalingen verschijnen bij bewegingswerkwoorden makkelijker na het werkwoord dan bij statische werkwoorden)¹⁵. Het onderzochte corpusmateriaal is echter te beperkt om deze laatstgenoemde variabelen betrouwbaar te analyseren.

In De Meersman (1988) verklaart de *complexiteit* van o.a. een VzC opnieuw de SVx-volgorde of achterveldpositie. In tegenstelling tot het onderzoek van 1980 interfereert *complexiteit* wel met andere variabelen, zoals de *bijzinslengte* (bijzinnen met 3, 4, 5, 6 of meer zinsdelen worden apart bestudeerd) en de *informatieve waarde* van de constituent (indefiniete nominale objecten staan meer in extrapositie dan definitieve objecten).

¹⁵ Deze bevinding werpt ook nieuw licht op het onderzoek van Vanacker (1973, 1977). Aan de hand van bandopnames toonde Vanacker aan dat de SVx-volgorde (of achterveldpositie) vaak voorkomt in het West-Vlaams en in het Frans-Vlaams. Voor het Frans-Vlaams is het zelfs de meest gewone schikking in zinnen met een bewegingswerkwoord en een richtingsbepaling. Volgens Vanacker is deze achteraanplaatsing een oorspronkelijk Ingveoons verschijnsel waarbij ook de invloed van het Frans een rol kan spelen (1977: 213). De Meersman gelooft echter dat deze achteraanplaatsing een relict is van een vroeger taalstadium van het Nederlands met een relatieve plaatsingsvrijheid voor inherente bepalingen. Volgens de Meersman is het ook niet vreemd dat dit oude relict net in het West- en Frans-Vlaams bewaard is gebleven: “Het lijkt aannemelijk dat deze plaatsingsvrijheid, die in de meeste dialecten verloren is gegaan en in het A.N. niet terug te vinden is, precies in het perifere West- en Frans-Vlaamse gebied tot op zekere hoogte bewaard is gebleven” (de Meersman 1980: 120).

De Schutter (1988) merkt net zoals Coussé (2003) en Van der Horst (2008) op dat het in het Middelnederlands goed mogelijk was om twee of zelfs meer constituenten achter de werkwoordelijke groep te plaatsen en dat ook NC's en adjectieven zonder (veel) bezwaar in extrapositie konden staan. Hij duidt drie mogelijke verklaringen aan voor de achterveldpositie: 1. de *complexiteit*, 2. de *grammaticale status* en 3. de *pragmatische functie*.

De eerste verklaring, de *complexiteit* van een zin, werd ook al door De Meersman (1988) vastgesteld en past volgens De Schutter binnen de hedendaagse hypothese dat extrapositie de tangconstructie¹⁶ kan verlichten (cf. ANS 1997: 1374). Extrapositie verlicht deze spanning: “de hoorder krijgt het (hoofd)werkwoord wat vroeger te horen en kan dus ook vroeger met de definitieve ‘processing’ van de zinsbetekenis beginnen” (De Schutter 1995: 197). Volgens De Schutter is er sprake van een extrapositieregel die, in vergelijking met het hedendaagse Nederlands, meer dan één keer toegepast kan worden in het Middelnederlands (vandaar een recursief optredende optionele extrapositieregel). Tegelijkertijd komt een meervoudige toepassing van de extrapositieregel ook niet zo frequent voor (ter illustratie: in de 130 zinnen komt enkelvoudige extrapositie 62 keer voor, meervoudige maar 23 keer; De Schutter 1988: 390).

De *grammaticale status* van een zinsdeel kan eveneens de zinspositie verklaren. Zoals De Meersman (1980) opmerkte, zijn voornamelijk bijwoordelijke bepalingen in de vorm van een VzC vrij om in achterveldpositie (of SVx-volgorde) te verschijnen. Ook De Schutter toont aan dat deze VzC's (De Schutter spreekt over *specificeerders*), samen met de bijzinnen, de zinsdelen zijn waarbij extrapositie eerder de regel dan de uitzondering is (ter illustratie: 60% van de VzC-specificeerders staat in achterveldpositie en 100% van de bijzinnen). VzC-complementen, zoals voorzetselvoorwerpen en inherente bepalingen, die in syntactisch opzicht dichter bij het werkwoord aansluiten, komen in 33% van de gevallen in extrapositie voor. De Schutter vergelijkt deze resultaten met de verdeling van VzC's in het moderne Nederlands en merkt daarbij op dat voorzetselvoorwerpen in het moderne Nederlands veel frequenter in extrapositie verschijnen in vergelijking met het Middelnederlands (51,3% in achterveldpositie). Wat de VzC-specificeerders betreft, zien we een daling van 60% in het Middelnederlands naar gemiddeld 40%¹⁷ in het moderne Nederlands. Enkel de oorzaakbepalingen staan in het moderne Nederlands zeer frequent in het achterveld (71,6%). Op basis van deze gegevens besluit De Schutter dat niet zozeer de *grammaticale functie* extrapositie beïnvloedt in het moderne Nederlands, maar wel dat “de syntactische structuren zich naar de eisen van pragmatische onderscheidingen plooien” (De Schutter 1988: 398). Voorzetselobjecten en bepalingen van oorzaak hebben

¹⁶ Ook De Schutter houdt rekening met een subject-werkwoord-tang (cf. De Meersman 1980, 1988) en niet met een bindwoord-werkwoord-tang (cf. ANS).

¹⁷ Dit cijfer werd berekend door alle percentages achterveldpositie voor de verschillende specificeerders (*tijdstip*, *omstandigheid*, *oorzaak*, *middel/hoedanigheid*, *duur* en *richting/plaats*) samen te tellen en te delen door zes (het aantal types VzC-specificeerders).

namelijk een sterk verklarend karakter ten opzichte van de stand van zaken. Verklarende constituenten komen vanzelf in een topic-focus-spanningsveld terecht en zo'n topic-focus-oppositie zet extrapositie in gang (cf. De Schutter 1985). De *grammaticale functie* van zinsdelen speelt dus voornamelijk in het Middelnederlands een rol.

De *pragmatische functie* is volgens De Schutter (1988) de derde verklaring. Achterveldpositie hing in het (vroeg) Middelnederlands vrijwel exclusief met sterkere focaliteit samen, in tegenstelling tot het moderne Nederlands waar er volgens De Schutter een ruimere pragmatische functionaliteit heerst met een topic-focus-spanning. Aan de hand van een (beperkte) kwalitatieve studie toont De Schutter (1988) aan dat extrapositie van definiëte¹⁸ constituenten in het Middelnederlands gelieerd is met grotere prominentie en vrijwel altijd met sterke focaliteit. Een plaatsing voor het werkwoord is de regel wanneer de constituent een entiteit noemt die een stevige verankering in de co- of context heeft (De Schutter 1988: 397-398). Deze bevinding is volgens De Schutter ook in verband te brengen met het onderzoek van de Meersman (1980): “hoe sterker een constituent in focus komt te staan, hoe sterker ook de neiging wordt om die heel specifiek te beschrijven” (De Schutter 1988: 398) en dus hoe complexer de constituent zelf wordt (cf. De Meersman 1988).

De Schutter herhaalt zijn onderzoek in 1995 maar gebruikt rijmteksten als informatiebron. Aanleiding is Van den Berg (1991) die de bruikbaarheid van Middelnederlandse poëzie voor syntaxisonderzoek ter discussie stelt. De Schutter bespreekt dezelfde variabelen – *complexiteit*, *grammaticale status* en *pragmatische functie* – en stelt vast dat er geen noemenswaardige verschillen bestaan tussen de resultaten gebaseerd op proza en die op poëzie¹⁹. Wel verschillend is de operationalisatie van de pragmatische variabele. In het onderzoek van 1988 bestudeert De Schutter maar een fractie van het totale materiaal²⁰. In het onderzoek van 1995 bestudeert hij het volledige corpusmateriaal en krijgt iedere constituent een waarde: *topic* (“als de constituent iets noemt dat inherent deel uitmaakt van het onderwerp van de tekst, zoals die zich tot op dat ogenblik ontwikkeld heeft”), *focus* (“als een nieuw element of aspect aan het al gegeven complex wordt toegevoegd”) of *neutraal* (de restgroep) (De Schutter 1995: 201). Ondanks een verschil in materiaal, poëzieteksten versus proza, en een andere

¹⁸ Volgens De Schutter (1988) zijn indefiniëte constituenten sowieso sterk focalisch en hebben ze met andere woorden geen specifieke positie meer nodig. Enkel de definiëte constituenten zijn naar hun vorm niet duidelijk gedefinieerd, zodat een toepassing van de extrapositieregels als een expliciet teken voor focaliteit kan gelden.

¹⁹ De Schutter merkt wel op dat er een correlatie bestaat tussen het syntactische feit van de extrapositie en het poëtische van de rijmpositie (De Schutter 1995: 203).

²⁰ De Schutter (1988) operationaliseert de derde variabele niet voor het volledige corpus. Hij bespreekt slechts enkele zinnen met een definiëte constituent in extrapositie en bevestigt op basis van deze beperkte selectie dat vooral definiëte constituenten die zinsfocus dragen in extrapositie staan.

operationalisatie van de pragmatische variabele, blijft het uiteindelijke resultaat vergelijkbaar en staan vooral focusconstituenten in extrapositie.

In 2003 gaat De Schutter nogmaals in op de zinsbouw van het (vroeg) Middelnederlands en gebruikt hij de Gentse leprozerie uit 1236 als onderzoeksmateriaal. In deze studie wordt de pragmatische motivering om de focusconstituent achterin de zin te plaatsen opnieuw bevestigd, net zoals de rol van de zinscomplexiteit. Door een zinsdeel of een zinsdeelstuk (afkomstig van een semantisch zware constituent die in tweeën wordt gesplitst) in het achterveld te plaatsen, hoeft de hoorder/lezer niet te lang te wachten op het structurerende element, het hoofdwerkwoord, wat de interpreteerbaarheid ten goede komt (De Schutter 2003: 49). Verder speelt de *grammaticale functie* een rol en is er opnieuw een duidelijk verschil te merken tussen de verplaatsbaarheid van (nominale en prepositionele) complementen enerzijds en adjuncten anderzijds (cf. De Schutter 1988). Deze tendens neemt sterk af in het hedendaagse Nederlands: nominale complementen en adjuncten komen in het huidige Nederlands niet meer in aanmerking voor extrapositie. Deze evolutie is vanaf de 16^e eeuw ontstaan. Een mogelijke verklaring is dat in die tijd de laatste resten van het formele naamvalsysteem in nominale constituenten gesneuveld zijn (De Schutter 2003: 59-60).

Samengevat kunnen we besluiten dat de variabelen *complexiteit* en *informatiewaarde* volgens De Meersman en De Schutter een effect hebben op de achterveldpositie van VzC's. De Meersman (1980) heeft het daarbij in eerste instantie over de *complexiteit van de VzC*, De Meersman (1988) en De Schutter (1988, 1995, 2003) over de *complexiteit van de zin*, ook wel de afstand tussen het subject en het werkwoord (of het middenveld). De *informatiewaarde* of de *pragmatische functie* speelt volgens beide auteurs in het Middelnederlands een cruciale rol, in de mate zelfs dat alle zinsdelen met *zinsfocus* (in welke vorm en met welke grammaticale functie dan ook) in het achterveld konden staan. Andere variabelen die achterveldpositie in het Middelnederlands verklaren, zijn het *bijzinstype* (De Meersman 1980) en de *grammaticale functie van de VzC* (De Meersman 1980; De Schutter 1988).

Hoeksema (2014)²¹, tot slot, beperkt zijn onderzoek tot voorzetselvoorwerpen. Hij stelt vast dat voorzetselvoorwerpen vanaf 1800 veel frequenter in achterveldpositie staan in vergelijking met vroegere fases van het Nederlands (cf. De Schutter 1988). Dat is op zich verrassend aangezien achterveldpositie in het verleden veel vrijer was. Volgens Hoeksema “fluctueert ook de gemiddelde lengte van de voorzetselvoorwerpen mee met de mate van extrapositie, al lijkt daarmee niet de hele groei te kunnen worden verklaard” (Hoeksema 2014: 236).

²¹ Hoeksema onderzoekt voorts de factoren *lengte* (in lettergrepen) en *definietheid* van de voorzetselvoorwerpen maar gebruikt daarvoor enkel recenter corpusmateriaal (1950-2014). We bespreken deze factoren daarom bij de synchronoon taalgebruiksgebaseerde studies (§2.2.4).

2.2.1.3 Conclusie historisch onderzoek

In het Middelnederlands stonden VzC's als bijwoordelijke bepaling meer in het achterveld in vergelijking met het moderne Nederlands (60% versus 40%; De Schutter 1988) en voorzetselvoorwerpen misschien wel verrassend minder (33% versus 51%; De Schutter 1988). Ongeacht de procentuele verschillen bestaat de keuze tussen middenveld en achterveld al eeuwenlang. De motivatie om een VzC buiten de tang te plaatsen is wel veranderd in de tijd. Voornamelijk de rol van de *zinsfocus* is gewijzigd (cf. De Meersman 1988, De Schutter 1988, Burridge 1993, Coussé 2003), in die mate zelfs dat we van een zuiver discoursgemotiveerde verklaring naar een verklaring in termen van *syntactische complexiteit* kunnen spreken. In het Middelnederlands konden allerlei zinsdelen met focus in het achterveld staan. In het moderne Nederlands komen vooral complexere VzC's, die een nabepaling of een opeenstapeling van nabepalingen hebben (waardoor de VzC's ook langer zijn), in achterveldpositie voor. Qua opbouw is een VzC sowieso complexer dan bijvoorbeeld een nominale constituent, omdat hij uit twee delen bestaat: een *relator* (het voorzetsel) en *relatum* (wat er op het voorzetsel volgt). Het valt op dat constituenten met een complexere bouw, zoals VzC's maar ook voegwoordelijke constituenten en bijzinsconstituenten (alle bestaande uit een *relator* en *relatum*), hun positie in het achterveld bewaard hebben. Andere constituenten zijn vanaf de 16^{de} eeuw, en zeker in de 17^{de} eeuw, hun achterveldpositie verloren. Of dit het gevolg is van een veranderende onderliggende motivatie (van *discoursgemotiveerd* naar *syntactisch complex*), van een sneuvelend formeel naamvalsysteem (De Schutter 2003), dan wel van de opkomst van de eenheidstaal (met meer regels en normen), is nog grotendeels een open vraag. Hoe het ook moge zijn, bovenstaand overzicht illustreert dat er al heel wat interessant onderzoek naar dit taalfenomeen verricht is. Toch is er nog heel wat corpusonderzoek vereist, dat systematisch verzameld is en goed gedifferentieerd naar een aantal beïnvloedende variabelen. Dat is trouwens niet alleen van belang voor de historische taalkunde, maar evengoed voor de taalkunde die zich bezighoudt met de volgordevariatie vanuit een puur synchroon empirisch perspectief, zoals onderhavig onderzoek.

2.2.2 Generatief onderzoek²²

Hoewel het in de inleiding al duidelijk geworden is dat we in dit onderzoek opteren voor een taalgebruiksgebaseerde aanpak, willen we tegelijkertijd de rijke generatieve onderzoekstraditie niet links laten liggen. We bespreken eerst hoe het generatieve

²² Overeenkomstig de generatieve traditie opteren we ervoor om bij de beschrijving van de generatieve studies over de woordvolgordevariatie van VzC's de Engelstalige begrippen en afkortingen te gebruiken, bijvoorbeeld PP of prepositional phrase i.p.v. VzC en NP of nominal phrase i.p.v. NC.

grammaticamodel het taalgebruik als een afgeleide van een abstracte mentale representatie beschouwt. Vervolgens worden specifieke onderzoeken naar de (optionele en verplichte) extrapositie van o.a. VzC's besproken in drie opeenvolgende historische fases: 1. de constructiespecifieke opvatting (§2.2.2.1), 2. de principes- en-parameters-visie (§2.2.2.2) en 3. de minimalistische opvatting (§2.2.2.3). Tot slot volgt er een conclusie (§2.2.2.4) waarin we onze bedenkingen over deze formele aanpak toelichten en waarin we verdedigen waarom we in dit onderzoek voor een empirisch kwantitatief-corpusgebaseerde aanpak opteren.

Heel wat generatieve taalkundigen hebben zich beziggehouden met de woordvolgordevariatie van VzC's en met de syntactische regels in de Universele (mentale) Grammatica die de keuze tussen middenveldpositie of achterveldpositie kunnen beschrijven en verklaren. De standaardopvatting is dat elk kind geboren wordt met een Universele Grammatica, een eindige set syntactische regels die voor iedere moedertaalspreker van gelijk welke taal identiek is (Broekhuis 2001: 4). Op basis van deze regels en in combinatie met de input uit de omgeving slaagt een kind erin de grammatica van zijn eigen doeltaal te construeren. Deze verwerving verloopt in een aantal stadia: uitgaande van een beginstadium S_0 (de Universele Grammatica) doorloopt het kind een aantal tussenliggende stadia tot het uiteindelijk een (min of meer) stabiel eindstadium S_n (de doeltaal) bereikt (Chomsky 1995). De doelstelling van generatieve taalkundigen is om inzicht te krijgen in deze syntactische regels die an sich niet toegankelijk zijn en die bijgevolg afgeleid moeten worden uit alle mogelijke grammaticale zinnen. Sommige generatieve taalkundigen gebruiken daarvoor empirische gegevens (bijvoorbeeld Hoeksema 2008), anderen volgen Kerstens & Sturm (1979), die zeggen dat een taalkundige er intuïtief in slaagt een onderscheid te maken tussen grammaticale en ongrammaticale zinnen, omdat hij net zoals elke andere moedertaalspreker over een aangeboren taalvermogen beschikt dat enkel grammaticale zinnen genereert.

Een speciale testcase bij de zoektocht naar de syntactische regels zijn allerlei variationele fenomenen, en zo ook het verschijnsel extrapositie. Om de set regels zo economisch mogelijk te modelleren, wordt er vaak een onderscheid gemaakt tussen een dieptestructuur en oppervlaktestructuur²³. Syntactisch synonieme zinnen, die verschillen qua vorm maar niet in de betekenis, kunnen tot één dieptestructuur herleid worden. Of omgekeerd: syntactisch homonieme zinnen, die dezelfde vorm hebben maar een verschillende betekenis, tot meerdere dieptestructuren. De variabele volgorde van VzC's in het midden- en achterveld kan in die redenering als een oppervlakteverschijnsel beschouwd worden, het resultaat van een onderliggende dieptestructuur waarbij slechts één volgorde de basis is. De zoektocht naar de dieptestructuur en de regels die de

²³ In de minimalistische opvatting (zie §2.2.2.3) en meer bepaald vanaf Chomsky 1995 wordt er geen onderscheid meer gemaakt tussen een oppervlakte- en dieptestructuur.

woordvolgordevariatie verklaren, heeft de voorbije decennia veel aandacht gekregen en gaat hand in hand met verwoede pogingen om de onderliggende woordvolgordestructuur van het Nederlands te bepalen. Dat zou een structuur moeten zijn die het Nederlands met andere West-Germaanse talen gemeen heeft, en in het bijzonder met het Engels en Duits. In de generatieve theorievorming kunnen we een verschuiving vaststellen van een meer SOV-volgorde naar een meer SVO-volgorde (met in elke fase tegenstanders die de andere volgorde verdedigen).

De discussie over de onderliggende woordvolgordestructuur van het Nederlands (SOV of SVO) wordt eveneens beïnvloed door ontwikkelingen binnen de generatieve grammatica zelf. Volgens Broekhuis (2005) kunnen die ontwikkelingen gekarakteriseerd worden als de overgang van een constructiespecifieke benadering (zie §2.2.2.1) naar een benadering waarin de grammatica uit een verzameling algemene principes bestaat. In de constructiespecifieke benadering, of de traditionele generatieve opvatting, bestaat het computationele systeem van een grammatica (i.e. de syntaxis) uit een verzameling taalspecifieke herschrijfgeregels en taalspecifieke transformaties en kan de volgorde waarin de transformaties worden toegepast variëren. De verschillen tussen talen kunnen dus op meerdere manieren verklaard worden (door de herschrijfgeregels, de transformaties of de volgorde van de transformaties), wat van het taalverwervingsproces een complexe, misschien wel onmogelijke taak maakt. Vandaar dat men in de latere ontwikkelingen op zoek gaat naar algemene principes die het taalverwervingsproces aanzienlijk vereenvoudigen. De twee belangrijkste opvattingen zijn de principes-en-parameters-visie (§2.2.2.2) en de minimalistische opvatting (§2.2.2.3). Het grote verschil tussen beide is dat in deze laatste opvatting het computationele systeem niet langer *geparametriseerd* is. De verschillen tussen talen zijn in de minimalistische opvatting tot een minimum beperkt: de basisstructuur van elke taal is nagenoeg identiek en elke taal beschikt over dezelfde syntactische regels (Broekhuis 2001: 13). Enkel het moment waarop de noodzakelijke syntactische bewerkingen plaatsvinden, zorgt voor het verschil tussen natuurlijke talen. Volgens Broekhuis (2001: 13) is de taak van het taallerende kind bijgevolg gereduceerd “tot het achterhalen van de kenmerkspecificaties van de functionele hoofden en het vaststellen van de sterkte van de verschillende formele kenmerken”.

Het mag duidelijk zijn dat onderstaand overzicht op geen enkele wijze aanspraak wil maken op exhaustiviteit en dat het enkel een representatief beeld wil scheppen van hoe generatieve taalkundigen in de loop der jaren deze woordvolgordevariatie geïnterpreteerd en verklaard hebben.

2.2.2.1 De constructiespecifieke opvatting

Een start- en referentiepunt is Koster (1973, 1974), die het concept PP-over-V als een rechtse transformatie introduceert, conform de taalspecifieke transformaties in de traditioneel-generatieve opvatting (met traditioneel bedoelen we de generatieve visie uit de jaren '70). Het betekent eenvoudigweg dat de PP (of VzC) uit de basispositie voor het werkwoord omhooggetild wordt en na het werkwoord geplaatst wordt. Het werkwoord is het natuurlijke centrum of spiegelcentrum waardoor PPs links voor het werkwoord rechts in spiegelbeeld komen te staan (de zogenaamde spiegeltheorie). Daardoor staat een PP die nauwer verbonden is met het werkwoord (zoals PP₁) ook in het achterveld dichter bij V: ... PP_n ... PP_i PP_{i-1} ... PP₁ ... V ... PP₁ ... PP_{i-1} PP_i ... PP_n ...

Deze spiegeltheorie wordt ook door Wiers & Van Noort (1978) gebruikt in hun transformationele beregeling van bijwoordelijke bepalingen. Wiers & Van Noort weerleggen hiermee het eerdere idee van Booij (1973-1974) dat de onderlinge volgorde van de bepalingen onveranderd blijft wanneer deze postverbaal verschijnen. Dat Booij's theorie niet standhoudt, blijkt ook uit volgend voorbeeld (uit Florijn 1992: 170):

- (22) a Dat Jan zijn broer [tot mijn vreugde]_{PP2} [met een hamer]_{PP1} sloeg.
b Dat Jan zijn broer [tot mijn vreugde]_{PP2} sloeg [met een hamer]_{PP1}.
c *Dat Jan zijn broer sloeg [tot mijn vreugde]_{PP2} [met een hamer]_{PP1}.

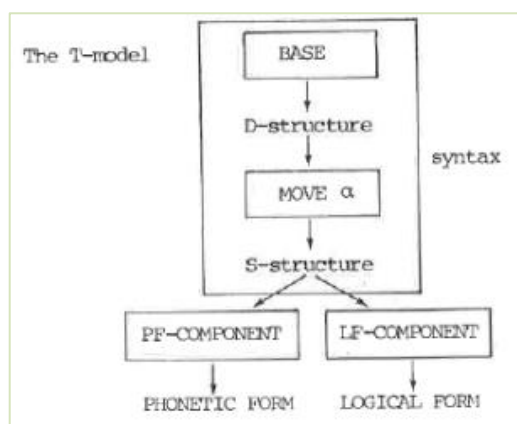
Koster gebruikt de spiegeltheorie om het Nederlands zowel in de diepte als aan de oppervlakte als een SOV-taal te beschouwen, in tegenstelling tot Bartsch & Vennemann (1972) en Kooij (1973) die uitgaan van een SVO-volgorde. De constituentenhiërarchie is in Koster's opvatting naar achteren gericht met als zwaartepunt het werkwoord. Doordat PPs gespiegeld kunnen worden tegenover dit werkwoord, lijkt het in sommige omstandigheden ten onterechte alsof de woordvolgorde SVO is. Bovendien is de SOV-volgorde de meest economische structuur om ook andere woordvolgordeverschijnselen in het Nederlands te verklaren, zoals de positie van een scheidbaar samengesteld werkwoord in een hoofd- en bijzin. In een SVO-taal zijn er twee regels nodig: de verplaatsing van het partikel naar achteren in de zin (*Particle Movement*) en de verplaatsing van het vervoegde werkwoord naar achteren in een bijzin (*Verb Final*). In een SOV-taal is er maar één regel nodig, namelijk een om het vervoegde werkwoord in een hoofdzin meer naar voren te verplaatsen (*Verb Placement*).

Koster motiveert de optionele rechtse transformatie aan de hand van het 'lekkende' karakter van de Nederlandse taal in vergelijking met een 'rigide' SOV-taal zoals het Japans (cf. Ross 1973). De PP-over-V-regel is dus een lekregel die de discontinuïteit in een zin opheft of verkleint. Stel dat een complexe PP in zijn basispositie staat (i.e. het middenveld) dan ontstaat er ongewenste discontinuïteit tussen de kern van de PP en V. Aangezien discontinuïteit geheugenbelasting met zich meebrengt die groter wordt

naarmate het tussenliggende materiaal complexer is (zie ook Bever 1970), verkleint de belasting als een complexe of lange PP achteraan staat (Koster 1974: 610).

2.2.2.2 De principes-en-parameters-visie

Onder andere Emonds (1976) *Structure Preservation Principle*²⁴ weerlegt het oorspronkelijke idee van een rechtse transformatie. Koster (1978) verklaart extrapositie daarom niet langer als een transformatie, een ‘move α ’, die de dieptestructuur tot een bepaalde oppervlaktestructuur transformeert, maar als een stilistische regel, die thuishoort in de fonologische component (PF) van de grammatica (onderaan links in het T-model; zie figuur 1 uit Hoekstra 1984: 22).



Figuur 1: T-model (uit Hoekstra 1984)

Volgens Koster bepalen idiosyncratische fonologische feiten welke elementen uit het middenveld geëxtraponeerd kunnen worden (Koster 1978: 55). Hoe deze feiten gedefinieerd moeten worden blijft echter onduidelijk. Koster verwijst enkel naar Guéron (1976) die stelt dat “extraposition from NP only applies under very limited conditions” (Koster 1978: 49). Zo kunnen bijvoorbeeld complementen enkel uit NPs geëxtraponeerd worden als die NPs focus²⁵ hebben en kan extrapositie tegengehouden worden door de aanwezigheid van een predicaat. Uit het voorgaande blijkt dat het hier om zeer onvoorspelbare en moeilijk te definiëren feiten gaat, die dan ook makkelijk weerlegd kunnen worden. Wel is het duidelijk dat Koster nog steeds vanuit een onderliggende SOV-structuur vertrekt waarbij de positie in het middenveld de basis is. Waarom complementen in de vorm van bijzinnen (CPs) verplicht postverbaal voorkomen

²⁴ “Emonds’ (1976) Structure Preservation Principle requires movement to target an independently motivated position” (Broekhuis & Corver 2016: 1593).

²⁵ “Guéron (1980) has argued on the basis of English that extraposition is possible only from noun phrases that are part of the focus (new information) of the clause [...]” (Broekhuis & Corver 2016: 1574).

(voorbeeld 23), kan Koster enkel verklaren door te veronderstellen dat in deze gevallen de postverbale positie vanuit de basis gegenereerd is.

- (23) a Peter heeft gezegd dat hij zou komen
b *Peter heeft dat hij zou komen gezegd

Hoewel De Haan (1979) weinig aandacht schenkt aan de extrapositie van PPs – hij vermeldt enkel dat afhankelijke PPs als nabepaling bij een nominale constituent²⁶ in extrapositie kunnen staan – bevestigt hij wel het idee van een basisgegenereerde postverbale positie en benoemt hij de verschillende extrapositievormen die hieraan beantwoorden: complementzinnen tegenover V (1979: 40-44), resultaatzinnen en type-II-relatiefzinnen uit NP (1979: 44-48). Bij basisgegenereerde ‘extrapositie’²⁷ is er geen sprake van een verplaatsing van een preverbale naar een postverbale positie, zoals dat aangenomen wordt in de traditionele opvatting, maar is de basispositie postverbaal. Er is echter één type zin (met als hoofd een NP) die volgens De Haan toch als transformationeel behandeld moet worden, namelijk type-I-relatiefzinnen die zowel in het middenveld (24a) als in achterveld (24b) kunnen verschijnen en die slechts naar één referent in de zin kunnen verwijzen.

- (24) a Hij heeft tegen de mensen [_S die aanwezig waren] gezegd [_S dat hij zou vertrekken].
b Hij heeft tegen de mensen – gezegd [_S die aanwezig waren] [_S dat hij zou vertrekken].

Type-I-relatiefzinnen kunnen we met afhankelijke PPs uit een NP vergelijken, indien deze PPs zowel in het midden- als in het achterveld kunnen staan. In dat geval is extrapositie bij beide types optioneel en verwijzen ze allebei naar één referent, de NP waarvan ze afhankelijk zijn²⁸. Vanuit deze veronderstelling is de PP(-uit NP)-over-V-regel een transformationele regel die zich in de syntaxis van een doeltaal afspeelt en niet in de fonologische component, zoals Koster (1978) beweert. Het idee van een stilistische regel

²⁶ “prepositional modifiers and complements of nominal phrases” (De Haan 1979: 44)

²⁷ Het woord *extrapositie* suggereert een verplaatsing terwijl er helemaal geen verplaatsing plaatsvindt. Dat is ook de reden waarom *extrapositie* tussen aanhalingstekens staat.

²⁸ De Haan (1974: 176-177) stelt dat extrapositie van PPs afhankelijk van (definiëte) indirecte objecten en subjecten uitgesloten is. Volgende voorbeelden (uit Broekhuis & Corver 2016: 1576) spreken echter voorgaande tegen:

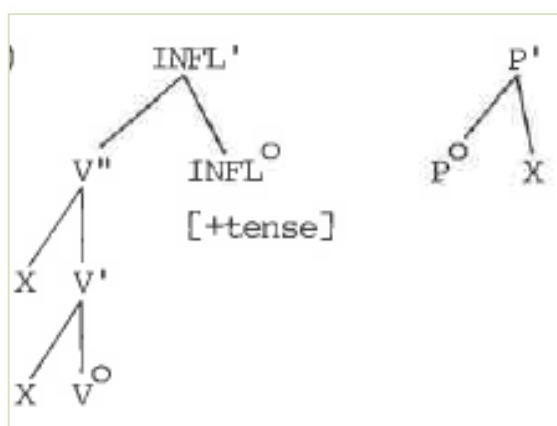
Marie heeft **veel mensen**_{IO} raad gegeven **met financiële problemen**.

Hier hebben altijd **veel mensen**_{SU} gewoond **met financiële problemen**.

Door de aanvaardbaarheid van bovenstaande voorbeelden komt ook de basisgegenereerde postverbale analyse onder vuur te staan. Deze analyse veronderstelt namelijk dat deze PPs in het achterveld kunnen staan als de ‘hoofden’ waarvan ze afhankelijk zijn ook basisgegenereerd zijn in de postverbale positie. “While this is plausible for objects, this is quite unlikely for subjects” (Broekhuis & Corver 2016: 1576).

wordt trouwens door meerdere taalkundigen tegengesproken, o.a. Koops (1984-85)²⁹; De Haan (1979), Hoekstra (1984) en Zwart (1990), die allemaal extrapositie als een syntactische regel opvatten. Ook Guéron (1980) “shows quite convincingly that extraposition must be a rule of syntax rather than a stylistic movement rule” (Hoekstra 1984: 74).

Een meer principiële verklaring waarom de postverbale positie in sommige gevallen verplicht (of basisgegenereerd) is, vindt Hoekstra (1984) bij Stowells *Case Resistance Principle* (1981). Het principe zegt dat enkel NPs in *case-gemarkeerde* posities kunnen verschijnen (en dat m.a.w. bijzinnen niet in de X-posities in figuur 2 kunnen staan, noch voor V, noch na P): “the requirement that NP’s must occur in Case-marked position, while clauses are not allowed in such positions” (Hoekstra 1984: 74).



Figuur 2: Overzicht Case-marked positions (Hoekstra 1984: 82)

Het *Case Resistance Principle* verklaart dus waarom de zinscomplementen in de voorbeelden 25, 26 en 27 (Hoekstra 1984: 82) niet voor V kunnen staan. Volgens Hoekstra zijn complementen basisgegenereerd in een postverbale positie (cf. Koster 1978, De Haan 1979) en hoeft deze postverbale positie niet verklaard te worden met behulp van woordvolgordecondities en specifieke transformaties (Hoekstra 1984: 94). Het *Case Resistance Principle* verklaart echter niet waarom de bijzinnen in 28a en 29a (Hoekstra

²⁹ Koops (1984-85) gelooft niet in een stilistische regel, noch in een formalistische aanpak om de transformatie aan de hand van condities, filters e.d. in het gareel te brengen (Koops 1984-85: 354). Postverbale PPs hebben volgens Koops hun eigen identiteit en bijgevolg hun eigen betekenis. Die verschillende identiteit verklaart waarom PPs die preverbaal een idiomatische lezing hebben in postverbale positie letterlijk genomen worden en waarom sommige postverbale PPs ongrammaticale of gewrongen zinnen opleveren, zoals de zin *dat Kuifje is in Afrika*. PPs die in de gemarkeerde postverbale positie voorkomen, hebben volgens Koops een incidenteel verband met het werkwoord. Ze zijn een modificatie van het werkwoord, ze beklemtonen de informatie, en moeten daarom als onafhankelijk geïnterpreteerd worden. Ze zijn bijgevolg geen loze vormvariant van een PP in preverbale positie (Koops 1984-85: 353). PPs die links van V staan hebben geen incidenteel verband met het werkwoord, waardoor andere factoren, zoals de intonatie en lexicale inhoud, dat verband moeten bepalen.

1984: 83), die makkelijk door een NP te vervangen zijn (zie 28b en 29b), wel in een *case-gemarkeerde* positie kunnen staan:

- (25) *dat [_{S'} dat Peter ook komt] mij ergerde
- (26) *dat Jan [_{S'} dat hij ziek geweest was] vertelde
- (27) *dat Jan over [_{S'} dat Piet ook mee ging] klaagde
- (28) a Voor [_{S'} dat de maaltijd begint] drinken we eerst iets
b Voor [_{NP} de maaltijd] drinken we eerst iets
- (29) a Door [_{S'} dat het zo heet was] konden we niets doen
b Door [_{NP} de hitte] konden we niets doen

Om de correcte zinnen 28a en 29a te verklaren, verwijst Hoekstra naar de *Unlike Category Constraint*, wat impliceert dat bijzinnen toch in een objectpositie kunnen staan (namelijk voor V). Volgens Broekhuis (2005) zijn Stowells (1981) *Case-Resistance Principle* en Hoekstras (1984) *Unlike Category Constraint* de meest invloedrijke antwoorden uit de jaren '80 op de vraag waarom verplichte extrapositie plaatsvindt. “Ze stipuleren eigenlijk beide dat zinscomplementen op bepaalde posities niet mogen voorkomen en derhalve moeten verplaatsen” (Broekhuis 2005: 232).

In tegenstelling tot De Haan komt extrapositie van PPs bij Hoekstra (1984) wel uitgebreid aan bod. Hoekstra verklaart aan de hand van de *Small-Clause Analysis* en de unidirectionaliteit op *regeren* waarom sommige PPs niet in extrapositie kunnen staan (zin 30a), terwijl andere PPs optioneel in extrapositie kunnen verschijnen (zin 31a en b).

- (30) a *dat Jan is in de tuin
b dat Jan in de tuin is
- (31) a dat Jan wandelt in de tuin
b dat Jan in de tuin wandelt

In zin 30 moet [_{PP}Jan in de tuin] als een PP-zin of *small-clause* beschouwd worden. De onderliggende structuur is dan: [_{V'} [_{NP} e] [_{PP}Jan in de tuin] is].

Hoekstras verklaring is gebaseerd op het *thèta-criterium*, een principe dat veronderstelt dat zinsdelen een *thèta-rol* van een *governed head* krijgen. Aangezien *dat Jan is ongrammaticaal is, kan het werkwoord niet *head governed* zijn, waardoor de NP Jan geen *thèta-rol* van dit werkwoord kan krijgen. Daarom besluit Hoekstra dat de *small clause* [_{PP}Jan in de tuin] een *thèta-rol* aan de NP Jan geeft. Een postverbale verplaatsing van deze PP-zin zou door de unidirectionaliteit op *regeren* impliceren dat niet alleen de PP in de tuin maar ook de NP Jan verplaatst moet worden. Omdat het werkwoord wel *case* geeft aan de NP Jan (Jan is het subject bij zijn) en deze NP dus enkel in een *case-gemarkeerde* positie kan staan (i.e. voor V, zie ook figuur 2), kan de volledige PP-zin niet verplaatst worden en dus ook de PP in de tuin niet. Het werkwoord wandelen in zin 31 geeft daarentegen wel een *thèta-rol* en *case* aan Jan, want de zin dat Jan wandelt is grammaticaal en Jan is subject bij het werkwoord wandelen. Het subject van de PP-zin is daardoor geen *trace*, noch een lexicale

uitdrukking: “the subject position is ungoverned and therefore PRO” (Hoekstra 1984:237). Kortom, het subject van de PP-zin is leeg en de onderliggende structuur is: [_V [_{NP}Jan]] [_{PP}PRO in de tuin] wandelt]. Volgens de unidirectionaliteit op *regeren* kan alleen de volledige PP-zin verplaatst worden, wat in zin 31a probleemloos gebeurt.

Kort samengevat verklaren de *Small-Clause Analysis* en de unidirectionaliteit op *regeren* waarom sommige PPs wel, en andere niet in extrapositie kunnen staan. Deze verklaring veronderstelt nog steeds dat het middenveld de basispositie is en dat het Nederlands bijgevolg een onderliggende SOV-volgorde heeft.

Net zoals Hoekstra gebruikt ook Zwart (1990) het *thèta-criterium* om te verklaren waarom afhankelijke PP-adjuncten wel geëxtraponeerd kunnen worden uit een NP in tegenstelling tot afhankelijke PP-complementen. In het volgende voorbeeld is *aan Marie* een complement en *over dat roken* een adjunct:

- (32) a dat ik die belofte aan Marie_{COMPL} over dat roken_{ADJ} geschonden heb
 b dat ik die belofte aan Marie_{COMPL} geschonden heb over dat roken_{ADJ}
 c *dat ik die belofte over dat roken_{ADJ} geschonden heb aan Marie_{COMPL}

Zwart maakt een onderscheid tussen voorzetsels die een *thèta-rol* aan de *noun* geven en voorzetsels die dat niet doen (cf. Emonds 1985: ‘nouns can assign theta roles only indirectly, via a preposition’). Complement PPs zijn afhankelijk van de *noun*, want ze geven de NP waarvan ze deel uitmaken een *thèta-rol*. Een positie in het achterveld van deze PPs zou betekenen dat er een *trace* aanwezig moet zijn in de NP en een link tussen de PP en deze *trace* (zodat de PP de *thèta-rol* kan overdragen). Voorts moet de *trace* aan de vereisten van het *Empty Category Principle* voldoen wat simpelweg betekent dat de *noun head governed* moet zijn. Volgens Zwart, in navolging van Rizzi (1989), is dat niet het geval, wat de ongrammaticaliteit van deze PPs in extrapositie verklaart. Adjunct PPs zijn onafhankelijk van de *noun* en kunnen daardoor in het achterveld geplaatst worden. Hetzelfde resultaat verkrijgt men in een *derivational approach* op basis van het *Projection Principle* (cf. Hoekstra 1987). Dit principe veronderstelt aanwezige *traces* bij een verplaatsing van complement PPs, terwijl *traces* van adjunct PPs zonder probleem gewist kunnen worden.

2.2.2.3 De minimalistische opvatting

In de minimalistische opvatting wordt de grammatica van een doeltaal sterk vereenvoudigd waardoor heel wat condities en principes afgeschaft en vervangen kunnen worden door één overkoepelend principe, het algemene principe *Full Interpretation*. Kaan (1993) gebruikt het *Full Interpretation*-principe om te verklaren waarom extrapositie van sommige afhankelijke PPs ongrammaticaal is. Ze vertrekt vanuit het idee dat afhankelijke PPs basisgegenereerd zijn in de rechtse perifere positie en baseert zich daarvoor op Koster (1978), Culicover en Rochemont (1990) en Rochemont en

Culicover (1990)³⁰. Kaan (1993: 146) veronderstelt dat vóór het moment of op het moment van de logische component (LF) de PP gelinkt moet zijn aan de NP waarvan de PP afhankelijk is. Ontbreekt deze link, dan wordt het *Full Interpretation*-principe geschonden en is de zin met een PP in het achterveld ongrammaticaal. Voorts motiveert Kaan op basis van het minimalistische framework van Chomsky (1992) waarom het idee van een rechtse verplaatsing achterhaald is³¹ (cf. Baltin 1981, 1983 en 1984). Extrapositie of eender welke verplaatsing is volgens het minimalistische programma enkel mogelijk als deze verplaatsing daadwerkelijk nodig is (i.e. *the concept of economy*), wat volgens Kaan niet het geval is bij een verplaatsing van een PP uit een NP. Bovendien stelt Kayne (1992) dat objecten enkel naar links verplaatst kunnen worden om de kenmerken van deze objecten te ‘checken’.

Broekhuis (2001) haalt nog twee vereenvoudigingen in de minimalistische opvatting aan. Ten eerste is het mogelijk om de oppervlaktestructuur af te schaffen. Dat representatieniveau diende voornamelijk “om te stipuleren dat een bepaald principe bevredigd moet worden voordat de fonologische kenmerken ingevoerd worden in de fonologische component, m.a.w. dat verplaatsing *overt* (‘zichtbaar’) plaats moet vinden” (Broekhuis 2001: 12). Nu maakt men een onderscheid tussen sterke kenmerken, die zichtbare verplaatsingen afdwingen, en zwakke kenmerken, die onzichtbare verplaatsingen afdwingen. Ten tweede wordt de syntactische component volledig bepaald door de Universele Grammatica. Syntactische verschillen tussen talen ontstaan dus niet doordat het computationele systeem andere herschrijfgeregels, condities op regeltoepassing of andere principes bevat (zoals in de principes-en-parameters-visie), maar doordat de kenmerken van de functionele hoofden verschillen vertonen in sterkte (Broekhuis 2001: 12). Dat betekent dat verschillen tussen talen volledig gereduceerd kunnen worden tot het lexicon.

³⁰ De aanname dat geëxtraponeerde constituenten in de rechterperiferie basisgegenereerd worden, wordt door De Vries (2009) tegengesproken. Bovendien merkt ook de *Syntax of Dutch* (Broekhuis & Corver 2016) op dat er problemen zijn met deze theorie (ook wel bekend als de *raising of promotion analysis*). Als afhankelijke PPs rechts van het werkwoord basisgegenereerd zijn, dan moet ook de NP in zijn basispositie rechts van de werkwoordelijke eindgroep staan. Bij objecten zou dat nog kunnen, vanuit een onderliggende SVO-volgorde, maar voor subjecten wordt dat onmogelijk (tenzij de hele basisvolgorde voor het Nederlands veranderd wordt, wat niet zonder enig gevaar is (Broekhuis & Corver 2016: 1576)). Verder merkt de *Syntax of Dutch* op dat afhankelijke PPs uit een NP geëxtraponeerd kunnen worden, terwijl de NP zelf ingebed is in een ander zinsdeel (bijvoorbeeld opnieuw een PP), bv. dat Jan [een boek [met foto’s [van zijn hond]]] heeft. “Although it is unclear to us what determines whether extraposition of a more deeply embedded PP leads to a generally accepted result or not, we conjecture that the restrictions are not of a syntactic nature, but that considerations of processing, semantic coherence, prosody, etc. are involved” (Broekhuis & Corver 2016: 1578).

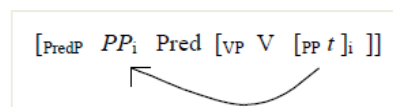
³¹ Zie ook De Vries (2009) voor meer informatie waarom een verplaatsing naar rechts problematisch is.

De komst van Kaynes *universal base-theorie* of *antisymmetrie-theorie* (1994) betekent een keerpunt binnen de generatieve grammatica. Alle talen zouden dezelfde onderliggende SVO-volgorde hebben, wat betekent dat OV-talen het resultaat zijn van een “leftward movement of various types of VP-internal constituents: DP-objects, PP-objects, predicative phrases, etc.” (Broekhuis 2008: 76). Het idee dat het Nederlands een onderliggende SVO-volgorde heeft, vinden we ook al terug bij Vanden Wyngaerd (1989) en Zwart (1992). Zij besluiten dat objecten in een Nederlandse SOV-structuur niet in hun basispositie staan. Deze SVO-volgorde en het idee dat er enkel linkse verplaatsingen zijn, zullen de basis vormen voor Koster's interpretatie van extrapositie in termen van een parallelle constructie. Natuurlijk zijn er ook tegenstanders: Haider (1997, 2000) en Barbiers (2000) houden vast aan een onderliggende SOV-structuur omdat deze volgorde een duidelijker inzicht geeft in de Nederlandse syntaxis³². Deze onderliggende structuur is ook de basis voor Barbiers *VP-Intrapositietheorie*. Hoewel deze theorie voor vele generatieve taalkundigen dissident is, is ze daarom niet minder interessant.

De parallelle constructie

Gemotiveerd door Kaynes *universal base-theorie* (1994) vertrekt Koster (1994, 1999b, 2000a, 2001b) vanuit het idee dat [_{VP} V PP] de onderliggende volgorde is en dat er geen rechtse verplaatsingen bestaan³³. In deze volgorde is extrapositie niet langer de uitzondering, maar de regel. Wat wel geëxpliciteerd moet worden is waarom sommige PPs verplicht en andere optioneel voor V staan.

Een oplossing vindt Koster in een *parallel construal*, een “larger set of structures in which two (or more) elements are juxtaposed and in which the second phrase specifies the first” (Broekhuis & Corver 2016: 1584), en in het onderscheid tussen zwakke en sterke PPs. Het uitgangspunt is dat alle complementen van V – of de zwakke PPs – verplicht naar links verplaatst moeten worden “to be functionally licensed” (Koster 2001b: 12; zie figuur 3).

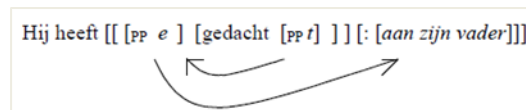


Figuur 3: Verplichte linkse verplaatsing van complementen volgens Koster (2001b)

³² In de SOV-volgorde heb je maar één regel nodig om andere taalverschijnselen te verklaren, namelijk *Verb Placement*.

³³ Waarom het Engels meer SVO oogt en het Nederlands meer SOV kan door een verschillende operationalisatie van *Object Shift* verklaard worden, wat Koster samenvat in een *Pied Piping Parameter*: “Dutch checks its VP-internal constituents individually by moving them separately, English collectively by moving the whole VP” (Koster 1999b: 8).

Sommige complementen links van V kunnen door de verplichte aanwezigheid van een target verplaatst worden – we merken hier echter meteen op dat het niet duidelijk is wanneer de aanwezigheid van een target verplicht is. De geëxtraponeerde PPs zijn een parallelle specificatie van een target in het middenveld. Het dubbelpunt of *colon* – vandaar ook de *Colon Phrase* – illustreert dat (zie ook figuur 4). Het target *[e]* stelt het echte complement in een positie links van V voor. Het is meestal een lexicaal element, maar het kan ook leeg zijn zoals in de volgende zin: *Ik heb (*het) gezegd dat hij zou komen*.



Figuur 4: Parallelle specificatie van het target *[e]* volgens Koster (2001b)

Het voorgaande spreekt Kaynes verbod op rechtse verplaatsingen (1994) niet tegen want “it is possible to expand standard syntactic structures to the right with asyndetic specifications” (Koster 1999c: 4). Omdat inherente PPs een vaste lexicale uitdrukking met het werkwoord vormen, hebben ze onvoldoende capaciteit om als een parallelle specificatie van een leeg complement te dienen³⁴.

Een aparte categorie zijn de onafhankelijke argumenten van V of adverbiale PPs, de zogenaamde sterke PPs, die volgens Koster zonder uitzondering aan de beide kanten van het werkwoord kunnen staan (Koster 2001b: 14):

- (33) a Hij heeft [tijdens de pauze] Marie gezien.

³⁴ Van Canegem-Ardijns (2006) waarschuwt voor een circulaire redenering. Zo is het volgens haar niet duidelijk waarom sommige PPs sterk zijn en andere zwak, zeker niet als ze beide semantisch verbonden zijn met het werkwoord (zie de volgende voorbeelden uit Van Canegem-Ardijns 2006: 433-434).

dat het er krioelde van de mieren
dat hij naar huis ging

Van Canegem-Ardijns besluit dat *van de mieren* in de eerste zin een sterke PP is en *naar huis* in de tweede zin een zwakke PP, simpelweg omdat de eerste niet geïncorporeerd kan worden in het middenveld en de tweede geïncorporeerd moet worden. “As such the distinction between strong and weak PPs turns out to be a circular one” (Van Canegem-Ardijns 2006: 434).

In lijn met Koster (2001b) zijn echter beide PPs zwak, aangezien ze allebei door het werkwoord opgeroepen worden. Volgens Koster's redenering is er in de eerste zin een target aanwezig waardoor *van de mieren* in extrapositie kan komen te staan en is *naar huis* in de tweede zin inherent verbonden met het werkwoord. In deze laatste zin kan *naar huis* geen parallelle specificatie van een leeg target vormen en is de VzC bijgevolg niet verplaatsbaar. Hoewel de aan-/afwezigheid van een target (en dus een parallelle specificatie) ongespecificeerd is, zou de grammaticale functie van de VzC een verklaring kunnen bieden. In de eerste zin gaat het over een voorzetselvoorwerp, net zoals in het voorbeeld ‘denken aan’ in Koster (2001b), en in de tweede zin over een complement. Bij voorzetselvoorwerpen kan een target voor een eventuele verplaatsing naar rechts zorgen, bij inherente complementen en niet-werkwoordelijke delen niet.

b Hij heeft Marie gezien [tijdens de pauze].

Ook in zin 33b veronderstelt Koster dat de rechtse positie van *tijdens de pauze* een specificatie is van een leeg target (visueel voorgesteld in zin 33c).

(33) c Hij heeft [_{PP} e] Marie gezien tijdens de pauze.

De sterke PP's staan in een preverbale positie hoger in de boomstructuur dan de zwakke PP's, wat het oorspronkelijke idee van de spiegeltheorie bevestigt: [PP₂ [[PP₁ [_{VP} V t₁]] [: PP₁]] [: PP₂]]

Extrapositie van afhankelijke PP's (ook wel *split extraposition*) wordt ook aan de hand van een *parallel construal* verklaard. Volgens Koster (2000b: 22) ziet de structuur van een afhankelijke PP (*uit India* in voorbeeld 34) er als volgt uit:

(34) [_{NP} [_{NP} een man] [: [_{PP} uit India]]] of [[_{XP} ... XP ...] &: XP]³⁵

De Vries (2011) gaat verder op de ingeslagen weg om zo de optionele extrapositie van afhankelijke betrekkelijke bijzinnen en voorzetselgroepen te verklaren (resp. zin 35 en 36).

(35) a Joop heeft een boek dat over taalkunde gaat gelezen.

b Joop heeft een boek gelezen dat over taalkunde gaat.

(36) a Joop heeft een boek over taalkunde gelezen.

b Joop heeft een boek gelezen over taalkunde.

In navolging van Koster (2000b) bespreekt De Vries extrapositie in termen van een specificerende nevenschikking. De geëxtraponeerde constituent wordt basisgegenereerd in een tweede conjunct, "dat op elk gewenst hiërarchisch niveau kan worden aangehecht, afhankelijk van de positie van het antecedent of anker" (De Vries 2011: 274). Een extra aanname is dat deze nevenschikking categoriaal gebalanceerd moet zijn, wat betekent dat beide conjuncten dezelfde syntactische categorie delen. "Het specificerende, tweede conjunct herhaalt syntactisch alles wat in het eerste element staat, maar dan met toevoeging van de betreffende modificeerder in zijn canonieke positie" (De Vries 2011: 286). Zinnen 37 en 38 illustreren het voorgaande: de beide conjuncten zijn functioneel gelijkwaardig want het zijn beide werkwoordelijke groepen.

(37) Joop heeft [&:P [een boek gelezen] [&: [~~een boek~~ dat over taalkunde gaat gelezen]]].

(38) Joop heeft [&:P [een boek gelezen] [&: [~~een boek~~ over taalkunde gelezen]]].

³⁵ Volgens de *Syntax of Dutch* roept dit voorstel een aantal vragen op: "the fact that the two conjuncts are not parallel in categorial status, syntactic function and meaning" (Broekhuis & Corver 2016: 1584)

Een verrassende consequentie is dat een deel van de kernzin fonologisch niet gerealiseerd wordt. Volgens De Vries is extrapositie daardoor een ellipsisconstructie (De Vries 2011: 291).

VP-Intrapositie

De *VP-Intrapositietheorie* van Barbiers vereist geen enkele rechtse verplaatsing, wat in overeenstemming is met Kayne (1994), maar verzet zich tegen de andere aanname, namelijk het idee van een universele SVO-structuur. De *VP-Intrapositietheorie* vertrekt met andere woorden nog steeds vanuit een onderliggende [_{VP} PP V]-structuur. Deze theorie kijkt ook op een ander punt af van alle hierboven vermelde theorieën en verklaringen. Barbiers beargumenteert immers dat de volgorde [_{VP} V PP] uit de vooropplaatsing van een (uitgebreide) projectie van het werkwoord resulteert (zie voorbeelden 39 en 40) en niet uit de verplaatsing van de PP zelf. Dat is op zijn minst opvallend, aangezien de werkwoordelijke eindgroep zo goed als altijd als een vaste, niet verplaatsbare pool wordt beschouwd.

(39) omdat hij ... [_{PP} in de tuin] [_{VP} een boek leest]

(40) omdat hij [_{VP} een boek leest] [_{PP} in de tuin] [_{VP} ~~een boek leest~~]

Een VP-Intrapositie ontstaat wanneer de geëxtraponeerde PP over de geïntraponeerde VP prediceert. Dat betekent dat de PP een predicaat³⁶ moet zijn van de VP en dat de VP dit predicaat als zijn subject neemt (Barbiers 2008: 170). Er zijn drie restricties: 1. argumenten kunnen niet optreden als predicaat, 2. een predicaat kan maar één subject hebben en 3. argumenten kunnen optreden als subject, predicaten kunnen dat niet. Deze restricties verklaren waarom VP-Intrapositie in sommige gevallen onmogelijk is. Verder stelt het *principe van Semantische Interpretatie*³⁷ dat VP-Intrapositie verplicht is in alle gevallen waar het mogelijk is. Omdat er makkelijk tegenvoorbeelden te vinden zijn, is er een extra regel nodig: “op het niveau van de fonologische interpretatie wordt de keuze gemaakt welke positie fonetisch wordt gerealiseerd: de basispositie van de verplaatste constituent of zijn landingsplaats” (Barbiers 2008: 171). Dat betekent tegelijkertijd dat de twee varianten geen gevolgen hebben voor de interpretatie op het semantische niveau want dat laatste werkt met de [_{VP} PP VP]-structuur in de syntaxis. Optie 1 en optie 2 uit figuur 5 betekenen bijgevolg hetzelfde:

³⁶ “De notie predicatie wordt hier uitgebreid van de traditionele relatie tussen een werkwoord en zijn subject naar relaties tussen andere constituenten zoals een bepaling en een werkwoordelijke constituent.” (Barbiers 2008: 5)

³⁷ Voor meer informatie, zie Barbiers (1995)

<p>Syntaxis: PP VP \Rightarrow VP PP VP (VP-Intrapositie)</p> <p>PF, optie 1: Spel de eerste kopie uit \rightarrow VP PP VP (geslapen in de tuin)</p> <p>PF, optie 2: Spel de laatste kopie uit \rightarrow VP PP VP (in de tuin geslapen)</p>

Figuur 5: Syntaxis volgens Barbiers VP-Intrapositie (Barbiers 2008: 172)

2.2.2.4 Conclusie generatief onderzoek

Het voorgaande overzicht maakt duidelijk dat de verklaringen voor het fenomeen extrapositie niet los te zien zijn van de historische ontwikkelingen binnen de generatieve grammatica. In de traditionele opvatting lag de focus op taalspecifieke herschrijfgeregels en transformaties, waardoor ook extrapositie als een rechtse transformatie werd geïnterpreteerd (Koster 1973 en 1974). Vanaf eind jaren '70 – begin jaren '80 verandert de focus en krijgen we achtereenvolgens de principes-en-parameters-visie en de minimalistische opvatting, die logischerwijs de theorieën over extrapositie beïnvloeden. Binnen de principes-en-parameters-visie vertrekken de theorieën vanuit een onderliggende SOV-structuur en veronderstellen ze allemaal dat verplichte extrapositiefenomenen (i.e. zinsdelen of zinsdeelstukken die verplicht in extrapositie staan) basisgegenereerd zijn in de postverbale positie. Toch is er ook een opvallend verschil waar te nemen. Koster (1978) beschrijft optionele extrapositie als een stilistische regel, terwijl alle andere taalkundigen het tot de syntaxis van het Nederlands rekenen (De Haan 1979, Hoekstra 1984, Zwart 1990). Daarnaast wijken ook de verklaringen en gebruikte principes in mindere of meerdere mate van elkaar af. De minimalistische opvatting ten slotte legt nog meer focus op een eenvoudig, economisch en algemeen grammaticaal model en dat zorgt voor nieuwe ideeën over extrapositie, in het bijzonder de parallelle constructie van Koster en de VP-Intrapositie van Barbiers. Deze theorieën lijken op elkaar aangezien ze beide een rechtse verplaatsing uitsluiten (cf. Kayne 1994) maar verschillen wat de onderliggende volgorde van het Nederlands en het verplaatste element betreft. De parallelle constructie vertrekt namelijk vanuit een onderliggende SVO-structuur, terwijl Barbiers VP-Intrapositie vasthoudt aan een onderliggende SOV-volgorde. Bovendien verplaatst de (on)afhankelijke PP zich in de parallelle constructie, net zoals in de andere theorieën trouwens, wat niet het geval is bij Barbiers *VP-Intrapositietheorie*. Barbiers theorie veronderstelt een onderliggende [VP PP VP]-structuur, waar niet de PP maar de VP verplaatst wordt. De *VP-Intrapositietheorie* vertoont dan weer wel een gelijkenis met Koster (1978). Bij beide auteurs speelt de keuze tussen een pre- of postverbale positie zich namelijk in de fonologische component (PF) af.

Hoewel de gepresenteerde analyses en onderzoeken in bovenstaand overzicht een interessant licht werpen op dit taalfenomeen in de Universele Grammatica, zijn er toch enkele losse eindjes op te merken. Ten eerste houdt de generatieve taalkunde vast aan een basisstructuur en een afgeleide maar bestaat er (nog steeds) geen consensus over die zogenaamde 'basisvolgorde' (SOV of SVO). Ten tweede is het nog steeds onduidelijk wat

optionele extrapositie precies triggert en waarom in het taalsysteem twee of meer varianten blijkbaar nodig of aantrekkelijk zijn (cf. Jansen 1978: 72). Ten derde resulteren de besproken analyses vaak in zeer abstracte verklaringen die voldoen aan systeeminterne motivaties zoals economie. Toch lijkt een economische motivatie weinig betekenisvol als de verklaring niet gekoppeld wordt aan discursieve of cognitieve mechanismen. Sterker nog, zelfs al neemt men aan dat het concrete taalgebruik afgeleid kan worden op basis van een overkoepelend volgordeprincipe – wat nog een punt van discussie is – dan nog hebben de transformaties of verplaatsingen enkel een theoretische waarde en kunnen ze de variabele fenomenen in het taalgebruik onvoldoende verklaren (cf. Givón 1979: 5-6).

De generatieve benaderingswijzen zoals we ze in dit hoofdstuk besproken hebben, staan enigszins haaks op een empirische visie op volgordevariatie. In een empirische onderzoekstraditie ligt de klemtoon net op uiteenlopende variabelen die de distributie tussen de volgordevarianten in het concrete taalgebruik bepalen. In de inleiding werd al aangegeven dat de onderliggende motivaties voor middenveldpositie en achterveldpositie in deze studie centraal staan. Om deze onderzoeksvraag te beantwoorden, is er een aanpak nodig die een wisselwerking tussen het taalvermogen en het taalgebruik impliceert en die kwantitatieve gegevens en statistische technieken gebruikt. Een empirische, multivariate analyse beantwoordt o.i. wel aan deze vereisten.³⁸

2.2.3 Functioneel grammaticaal onderzoek

Ook buiten de generatieve taalkunde zijn er enkele grammaticale analyses over extrapositie ontwikkeld, meer bepaald twee studies binnen de functionele (discourse) grammatica (Vandenbosch 1985; Van de Velde 2012) en een binnen de constructiegrammatica (Van Canegem-Ardijns 2004, 2006). Deze studies worden in de volgende paragrafen kort voorgesteld.

2.2.3.1 Functionele (discourse) grammatica

Het onderzoek van Vandenbosch (1985) kadert binnen de functionele taalkunde. Vandenbosch betwist een puur syntactische verklaring om de variabele positie van VzC's te bespreken en stelt een hiërarchisch systeem voor dat uit een *pragmatische* en *syntactische component* bestaat. In de *pragmatische component* is een element, zoals een VzC,

³⁸ Vanuit de generatieve hoek wordt een empirische benadering dan weer afgedaan als een loutere beschrijving van de taalfeiten die niet bijdraagt tot een beter begrip van de organisatie van de taal op een mentaal abstract niveau (zie ook Newmeyer 2003a en Bybee 2006 voor een discussie over een formele versus een taalgebruiksgebaseerde visie op taal).

een abstract element en krijgt het wel of geen zinsfocus. Vervolgens komt het in de *syntactische component* terecht, waarbij ook syntactische beperkingen een rol kunnen spelen. Een inherente VzC die bijvoorbeeld in het pragmatische level zinsfocus krijgt, zal door syntactische beperkingen niet in extrapositie kunnen staan.

Met behulp van een corpusonderzoek stelt Vandenbosch (1985) vast dat er een subtiel pragmatisch verschil bestaat tussen VzC's in achterveldpositie. Enerzijds krijgt een VzC in extrapositie meestal zinsfocus, of 'directe focus'. De spreker kiest ervoor om het belangrijkste element te beklemtonen en plaatst het net daarom in de positie na de werkwoordelijke eindgroep. De focusstatus wordt ook duidelijk door een stijgend intonatiepatroon. Anderzijds kan extrapositie net gebruikt worden als 'indirecte focus'. Het gaat daarbij om zinnen met meerdere VzC's waarvan één de thematische is en de andere de zinsfocus draagt. Vandenbosch stelt vast dat de thematische VzC geëxtraponeerd wordt en dat de VzC met zinsfocus onmiddellijk aan de werkwoordelijke eindgroep voorafgaat. Deze laatste VzC krijgt 'indirect' focus door het contrast met de thematische informatie in extrapositie. Het contrast wordt verder bevestigd door een dalend intonatiepatroon. Op die manier verklaart Vandenbosch ook waarom sommige VzC's niet in aanmerking komen voor extrapositie. In een zin met bijvoorbeeld drie VzC's waarvan de eerste zinsfocus krijgt, de tweede thematische informatie bevat en de derde VzC noch zinsfocus heeft, noch thematische belangrijk is, is extrapositie van deze laatste VzC zeer onwaarschijnlijk. Extrapositie zou namelijk suggereren dat deze VzC of informatie in focus (directe focus) of thematisch belangrijke informatie (indirecte focus) presenteert, wat zeer verwarrend is voor de hoorder.

Van de Velde (2012) onderzoekt hoe extrapositie van VzC's behandeld wordt in de *Functional (Discourse) Grammar*³⁹. Het gaat meer bepaald over extrapositie van afhankelijke VzC's uit een NC zoals in voorbeeld 41.

(41) Ik heb twee boeken gelezen over Amsterdam.

Extrapositie van de VzC *over Amsterdam* zorgt voor een discontinuïteit: "gevallen waarin beide delen ondanks hun fysieke scheiding toch één constituent vormen" (Van de Velde 2009: 57). Dergelijke gevallen zijn een uitdaging voor de formele syntaxis. Voegen we haakjes toe aan voorbeeldzin 41 dan zijn er twee mogelijkheden (42a en b).

(42) a [[twee boeken]_{NP} gelezen [over Amsterdam]_{PP}]_{VP}
 b [twee boeken [gelezen] [over Amsterdam]_{PP}]_{NP}

Elk op hun beurt zijn ze problematisch: sluit het haakje na *boeken* dan maakt de VzC *over Amsterdam* geen deel uit van de NC; sluit het haakje na *Amsterdam* dan maakt *gelezen* deel uit van de NC. Binnen de generatieve grammatica lost men dit op door twee structuren te

³⁹ Voor een introductie tot de *Functional Discourse Grammar*, zie Hengeveld & Mackenzie (2008)

gebruiken: één dieptestructuur waarin de VzC direct verbonden is met de NC en één oppervlaktestructuur waarin ze gescheiden zijn (zie o.a. Koster 1973, 1974 en 1978). *Functional Discourse Grammar* gebruikt echter geen transformaties: “the constituent tree has to spell out directly in its right form when it gets the input from the semantic and pragmatic representation” (Van de Velde 2012: 436). *Functional Discourse Grammar* erkent met andere woorden geen regels die elementen uit een origineel geïntegreerde syntactische structuur verplaatsen. Volgens Van de Velde (2012) zijn VzC’s als nabepaling syntactisch onafhankelijk op het *Morphosyntactic Level*. Op het *Representational Level* wordt de nabepaling semantisch gelinkt aan het nomen of de NC waartoe het behoort.

As will be supported by a series of arguments, the link between the PP and the NP is merely of a semantic nature, and consequently, the linking can be done at the Representational Level. Morphosyntactically, the ‘attributive’ PPs can be shown to behave independently. (Van de Velde 2012: 443)

Volgens Van de Velde (2012) kunnen verschillende variabelen verklaren waarom een VzC geëxtraponeerd wordt. Zo kan extrapositie het resultaat zijn van topic-focus-toewijzing op het *Interpersonal Level*: “Focus assignment at the Interpersonal Level triggers extraposition at the Morphosyntactic Level” (Van de Velde 2012: 446). Naast deze pragmatische variabele *focus* kan extrapositie ook het gevolg zijn van *syntactische complexiteit*.⁴⁰

2.2.3.2 Constructiegrammatica

Van Canegem-Ardijns (2006) bestudeert de positie van afhankelijke VzC’s opgeroepen door een adjectief (ter illustratie: *tevreden over*) en verklaart haar bevindingen binnen de theorieën van de constructiegrammatica⁴¹ (meer bepaald van de *Radical Construction Grammar*; zie ook Croft 2001). Ze vergelijkt de woordvolgorde van deze VzC’s met die van voornaamwoordelijke bijwoorden en stelt vast dat als de voornaamwoorden gesplitst worden door een adjectief (bv. *daar tevreden over*), de VzC bij dat adjectief meer in extrapositie komt te staan:

If the two parts of the pronominal adverb or an adjective are frequently separated by the adjective, the prepositional object of the adjective very frequently occurs in extraposition and vice versa. (Van Canegem-Ardijns 2006: 438)

In zin 43 (Van Canegem-Ardijns 2006: 437) wordt het voornaamwoordelijk bijwoord *erover* gesplitst en staat het adjectief tussen het eerste en tweede deel. Overeenkomstig

⁴⁰ Deze theorie lijkt goed op een multifactoriële analyse van variatiepatronen. De verschillende levels herbergen dan een of meerdere beïnvloedende factoren.

⁴¹ Zie Verhagen (2005) voor een introductie tot de constructiegrammatica voor het Nederlands.

deze constructie zal in een zin met een VzC bij hetzelfde adjectief extrapositie de voorkeur krijgen (geïllustreerd in zin 44).

(43) Ik weet dat hij er tevreden over is.

(44) Ik weet dat hij tevreden is over iets.

De redenering is gebaseerd op het *product-oriented* schema van Bybee (1985) en past binnen de zogenaamde *usage-based approach*. Als een spreker meermaals geconfronteerd wordt met een structuur zoals in zin 43, dan is de kans groot dat hij dezelfde structuur (en dus volgorde) bewaart voor andere constructies, zoals de positie van een VzC in zin 44:

If a speaker is frequently confronted with the word order adjective-preposition in split pronominal adverb constructions, he may be inclined to preserve this order in other constructions as well, which will enhance the use of extraposition constructions. (Van Canegem-Ardijns 2006: 443)

The model can also be called a usage-based model, in which the representation of grammatical units in a speakers's mind is not only determined by the structure of grammatical patterns, but also by properties of the use of utterances in communication. (Van Canegem-Ardijns 2006: 448)

Het idee dat de analogie in de andere richting zou werken – met name dat de positie van een VzC ten opzichte van een adjectief de (optionele) splitsing van voornaamwoordelijke bijwoorden zou bepalen – is volgens Van Canegem-Ardijns zeer onwaarschijnlijk. Taalgebruikers voelen aan wanneer een voornaamwoordelijk bijwoord gesplitst kan worden en wanneer niet, “it is often a matter of grammatical acceptability” (Van Canegem-Ardijns 2006: 443), terwijl de mogelijke posities van VzC's ten opzichte van het adjectief zonder veel uitzonderingen aanvaardbaar lijken. Toch zullen taalgebruikers een bepaalde voorkeur hebben waardoor bij sommige adjectieven de VzC systematisch vaker in extrapositie verschijnt.

De verklaring op basis van analogie veronderstelt geen syntactische relaties tussen de elementen van een constructie, wat aan de *Radical Construction Grammar*-opvatting beantwoordt: “the [analogy] schemas fit the RCG conception of the absence of syntactic relations (compare Croft 2001: 203)” (Van Canegem-Ardijns 2006: 452). Anders gezegd, een taalgebruiker heeft de syntactische relaties niet nodig om de analogie tussen de splitsing van de voornaamwoordelijke bijwoorden en extrapositie van VzC's te begrijpen. Voorts worden de adjectieven onderverdeeld naargelang hun splitsingsmogelijkheden (+s en -s) en hun extrapositiemogelijkheden (+e, -e). De twee types adjectieven die ontstaan, zijn [+s, +e] en [-s, -e], respectievelijk adjectieven die vaak tussen het voornaamwoordelijk bijwoord staan en waarbij de VzC meestal in extrapositie verschijnt en adjectieven waarbij het voornaamwoordelijke bijwoord niet gesplitst wordt (of gesplitst wordt maar

niet door het adjectief zelf) en waarbij de VzC zelden in extrapositie staat. Deze twee types moeten niet als scherpomlijnde categorieën beschouwd worden, maar meer als de uitersten van een continuüm.

2.2.4 Empirisch onderzoek

Zoals aangegeven in de inleiding en in de vorige hoofdstukken wil huidig onderzoek inzicht geven in de precieze gebruiksvoorwaarden van de twee mogelijke woordvolgordes van VzC's in een bijzin, namelijk middenveldpositie en achterveldpositie. We kiezen daarom voor een empirische corpusgebaseerde studie waarbij we de data met behulp van statistische technieken testen. Vooraleer we dieper in kunnen gaan op de dataselectie en de onderzochte variabelen die de woordvolgorde kunnen verklaren, stellen we eerst de relevante synchrone variabelen uit eerder empirisch onderzoek in meer detail voor. We zullen deze eerdere studies ook kritisch bespreken en relevante en openliggende onderzoeksvragen vermelden.

In de eerste paragraaf (§2.2.4.1) bespreken we het pionierswerk van Jansen (1978, 1979) en maken de vergelijking met het onderzoek van Braecke (1990). In paragraaf 2.2.4.2 bespreken we Jansen (2003), die de drie positiemogelijkheden (voorveld, middenveld en achterveld) van een constituent in een hoofdzin linkt aan de *sentence-processing* theorieën van Hawkins (1994) en Gibson (1998, 2000). Tot slot komt het onderzoek van Hoeksema (2014) aan bod (§2.2.4.3), dat zich tot de variabele positie van voorzetselvoorwerpen beperkt. Historisch onderzoek van De Schutter (1988) had namelijk al vastgesteld dat voorzetselvoorwerpen in het moderne Nederlands zeer frequent in het achterveld staan (zie ook §2.2.1.2).

2.2.4.1 Eerste onderzoeken

2.2.4.1.1 Jansen (1978, 1979)

Corpus: gesproken taalgebruik van 40 oudere respondenten uit Leiden (50 jaar of ouder)

Methode: gestuurd interview (20 à 25 minuten) bestaande uit:

- formele vragen over het leven en de levenswijze van de informanten;
- informele open vragen over speciale perioden uit het leven van de informanten en hun ideeën en houding ten opzichte van maatschappelijke ontwikkelingen.

Jansen (1978) gaat na in hoeverre achterveldpositie bepaald wordt door een aantal taalinterne en -externe variabelen. De taalinterne variabelen zijn: *de grammaticale functie van de VzC*, *de accentuering van de VzC*, *de bepaaldheid van de VzC*, *de complexiteit van de VzC*, *de aanwezigheid van andere zinsdelen in het middenveld* (simpelweg de lengte van het middenveld), *de aard van de zin* (hoofdzin versus bijzin) en *de aanwezigheid van een anafoor*

aan het begin van de volgende zin die verwijst naar een referent binnen de VzC. De taalexterne variabelen zijn: *de stijl* (formeel versus informeel), *de individuele variatie*, *de sekse* (man versus vrouw) en *klasse* (hogere klasse versus lagere klasse). Het effect van deze taalinterne en -externe variabelen wordt in meer detail in de volgende alinea's besproken.

Voor de eerste taalinterne variabele, *de grammaticale functie van de VzC*, test Jansen (1978) of VP-bepalingen (bijwoordelijke bepalingen van *richting*, sommige plaatsbepalingen en voorzetselvoorwerpen; gebaseerd op De Haan 1976) meer bij de werkwoordgroep horen en dus meer in middenveldpositie staan in vergelijking met alle andere (predicaats)bepalingen. De redenering is dat VP-bepalingen nauwer met het werkwoord verbonden zijn en bijgevolg de positie voor de werkwoordelijke eindgroep verkiezen (i.e. middenveldpositie). Voor de richtingsbepalingen (en sommige plaatsbepalingen) blijkt dat het geval te zijn, voor de voorzetselvoorwerpen niet. Voorzetselvoorwerpen staan niet alleen meer in extrapositie in vergelijking met de andere VP-bepalingen, ze staan ook meer in extrapositie in vergelijking met de predicaatsbepalingen (cf. De Schutter 1988). Jansen verklaart dit door de voorzetselvoorwerpen als een apart niveau binnen de VP-bepalingen te beschouwen, met een mogelijk andere interne structuur dan de VP-bepalingen. Jansen expliciteert echter niet hoe deze structuur er precies uitziet.

In navolging van Overdiep (1949) gaat Jansen (1978) na of de plaatsing van een VzC in achterveldpositie een aantrekkelijkere zinsmelodie oplevert. Jansen veronderstelt daarbij dat een VzC met een accent op de referent (meestal een NC) frequenter in het achterveld voorkomt om zo twee opeenvolgende accenten te vermijden (als een VzC met een accent in het middenveld staat, ontstaan er twee opeenvolgende accenten: één op de NC (van de VzC) en één op de werkwoordelijke eindgroep). De absolute en relatieve cijfers tonen echter geen voorkeur voor achterveldpositie bij VzC's met een accent. Jansen stelde namelijk vast dat 40% van VzC's met een accent in achterveldpositie staan en 60% in middenveldpositie. Jansen besluit dat er geen relatie waargenomen kan worden tussen het accent en de plaatsing van de VzC.

Jansen (1978) stelt ook vast dat onbepaalde VzC's meer in extrapositie staan dan bepaalde. Volgens Jansen is dat een "afspiegeling van de meer algemene tendens in natuurlijke talen om de zin te openen met elementen die wijzen naar dingen die aan spreker en hoorder bekend zijn, en de verwijzers naar onbekende zaken daar achter te zetten" (Jansen 1978: 82)⁴². Jansen houdt hierbij enkel rekening met bepaalde en onbepaalde VzC's, zonder te verduidelijken hoe hij deze variabele geoperationaliseerd

⁴² Jansen & Wijnands (2004) werken het Links-Rechtsprincipe (of LRP) nog meer uit. Dit principe zegt dat de volgorde van de zinsdelen bepaald wordt door het informatiegehalte van de zinsdelen, conform de regels in het schrijfonderwijs en in de Algemene Nederlandse Spraakkunst. Met behulp van een corpusonderzoek tonen ze aan dat allerlei variabelen, negen om precies te zijn, het LRP doorkruisen.

heeft. We kunnen enkel vermoeden dat hij rekening houdt met (on)bepaalde determinatoren en voornaamwoorden.

Jansen (1978) onderzocht ook op welke manier de complexiteit van de VzC een rol speelt bij de plaatsing van VzC's. Hij baseert zich daarvoor op Behaghel (1909: 139):

So bildet sich unbewußt in den Sprachen ein eigenartiges rhythmisches Gefühl, die Neigung, vom kürzeren zum längeren Glied überzugehen [...] was ich [...] als das Gesetz der wachsenden Glieder bezeichnen möchte.

Later wordt de formulering: "Von zwei Gliedern von verschiedenem Umfang steht das umfangreichere nach" (Behaghel 1930: 85). Jansen stelt vast dat de spreker significant vaker achterveldpositie verkiest bij complexere VzC's. Complexiteit wordt in Jansens studie gedefinieerd als VzC's die een (of meer) bijvoeglijke bijzin of bijvoeglijke VzC bevatten of VzC's die uit twee of meer nevenschikte elementen bestaan⁴³. Jansen toetst daarbij zowel de volledige achteraanplaatsing van complexe VzC's als de mogelijkheid om deze VzC's te splitsen. Het verschil tussen beide operationalisaties is niet significant, wat volgens Jansen betekent dat "sprekers terwijl ze hun spraak plannen, het vermogen hebben om een zodanige rangschikking te maken dat ze zo min mogelijk moeilijkheden voor zichzelf maken. Daarbij kunnen ze twee gelijkwaardige strategieën gebruiken: splitsing en algehele verplaatsing" (Jansen 1978: 87).

De andere variabele die we onder de noemer *syntactische complexiteit* kunnen groeperen, namelijk de aanwezigheid van andere zinsdelen in het middenveld, vindt zijn oorsprong in van Haeringens hypothese, dat "a speaker wants to say together elements that belong together, for example subject and finite verb" (Van Haeringen 1947: 3). Zo is er in het onderzoek van Jansen een sterke stijging van het aantal VzC's in extrapositie zodra er 2 zinsdelen en 2-4 lettergrepen in de tang staan versus respectievelijk 0-1 zinsdeel en 0-3 lettergrepen. Daarna neemt het stijgende effect af.

De aard van de zin (hoofdzin versus bijzin) beïnvloedt de positie van VzC's niet. De laatste taalinterne variabele, de aanwezigheid van een anafoor aan het begin van de volgende zin, speelt wel een rol. De resultaten tonen aan dat VzC's met een anafoor frequenter in achterveldpositie staan. Dit kan evengoed een gevolg zijn van de complexiteit. De spreker kan namelijk aan een complexe VzC denken (bijvoorbeeld een VzC met een betrekkelijke bijzin als nabepaling) en op het laatste moment beslissen om er twee aparte zinnen van te maken. In dat geval staat soms een niet-complexe VzC toch in het achterveld. Volgens Jansen zeggen de resultaten ook iets over de manier waarop een spreker zijn zinnen

⁴³ Jansen en Wijnands (2004: 33) tonen later nogmaals aan dat complexere zinsdelen de eerste of laatste zinsplaats verkiezen.

bouwt. Hij veronderstelt namelijk dat “tijdens het uitspreken van het middendeel van de zin vaststaat hoe (het eerste stuk van) de volgende zin er uit gaat zien”⁴⁴ (Jansen 1978: 89).

Jansen onderzocht ook de invloed van de formele versus informele stijl en veronderstelt dat sprekers in een informele situatie minder rekening houden met de norm. Omdat normatieve grammatici plaatsing in het middenveld als de meest gewone en de meest normale variant beschouwen, verwacht Jansen meer middenveldpositie in de formele taal en een grotere kans op achterveldpositie in de informele taal. Jansen bakent daarvoor zijn interview af. Aan het begin van het interview stelde hij zagezegde “vrij precieze en formeel gestelde vragen over hun leven en de levenswijze” (Jansen 1978: 90), waarbij “lange antwoorden niet gestimuleerd werden” (Jansen 1978: 90). Het antwoord op deze vragen wordt als formele stijl gelabeld. Vervolgens kwamen wat informelere vragen aan de beurt: “1) speciale perioden uit het leven van de informanten” en “2) attitudes van de informanten ten opzichte van de recente maatschappelijke ontwikkelingen” (Jansen 1978: 90). Het tweede deel van het interview leverde volgens Jansen “wat meer informele spraak op” en er werden “langere antwoorden gestimuleerd” (Jansen 1978: 90). Het antwoord op deze laatste twee vragen kregen het label informele taal. Tegen de verwachting in stelt Jansen een significante daling vast van het aantal VzC’s in achterveldpositie in een informele situatie (35.9% in extrapositie in de formele situatie versus 32.8% in een informele situatie). Om deze resultaten te verklaren, verwijst Jansen naar Labovs (psycho)linguïstische redenering waarin beweerd wordt dat een spreker in eerste instantie aan zichzelf denkt en dat een spreker de volgorde van de zinsdelen zorgvuldiger plant wanneer er meer tijd is (cf. Labov 1972). Zinsdelen die daarvoor in aanmerking komen, kunnen dan ook makkelijker in extrapositie verschijnen. Kent de spreker tijdsgebrek dan zal hij meer vasthouden aan een strakke volgorde, of een onderliggende norm, in dit geval aan middenveldpositie. Om de resultaten te linken aan Labovs redenering, moeten we wel veronderstellen dat een spreker meer tijdsdruk ervaart in een informele situatie dan in een formele. In het geval het om een informele discussie of een informeel gesprek zou gaan, lijkt dat logisch want er is tijdens een gesprek een voortdurende voortgestuwde gespreksdynamiek die tevens voor meer druk zorgt. In het interview van Jansen gaat het echter over open vragen. De gespreksdynamiek is dus zowel tijdens het formele gedeelte als tijdens het informele gedeelte zeer beperkt. Labovs redenering zou m. i. enkel gebruikt kunnen worden als het taalgebruik van de informanten tijdens een informeel gesprek (of informele discussie) werd opgenomen.

⁴⁴ Jansen baseert zich hiervoor ook op onderzoek van Ford & Holmes (1978). Zij tonen aan dat een spreker tijdens een gesprek het volgende deel plant. Ford en Holmes gaan daarmee in tegen het aloude idee dat de vorm al vaststaat nog voor de taalgebruiker begint te spreken.

Jansen onderzocht ook het effect van *de individuele variatie*, van *de sekse* (mannen versus vrouwen) en van *de klasse* (hogere klasse versus lagere klasse)⁴⁵. Voor de individuele variatie merkt Jansen op dat er meer variatie is in de formele stijl in vergelijking met de informele stijl. Hij verbindt dat opnieuw aan bovenstaande redenering van Labov, namelijk dat sprekers in een formele situatie minder tijdsdruk ervaren en bijgevolg meer tijd hebben om te variëren. Jansen stelt verder ook vast dat 1. vrouwen meer VzC's in het middenveld plaatsen in vergelijking met mannen en dat 2. de hogere klasse meer VzC's in het middenveld plaatst dan de lagere klasse. Jansen verbindt de resultaten met een zekere normgevoeligheid die hoger is bij vrouwen en de hogere klasse dan bij resp. mannen en de lagere klasse.

Hoewel Jansen verschillende variabelen test en deze variabelen probeert te verklaren binnen het productie- en planningsproces, kunnen we toch enkele beperkingen vaststellen. Ten eerste is het onderzoek monofactorieel. De verschillende taalinterne en -externe variabelen worden afzonderlijk onderzocht, waardoor het onduidelijk is hoe betrouwbaar en generaliseerbaar de resultaten zijn en hoe de verschillende variabelen zich tot elkaar verhouden in hun effect op deze woordvolgordevariatie. Ten tweede baseren de resultaten zich op het taalgebruik van oudere participanten (50 jaar of ouder), wat vanzelfsprekend een effect kan hebben op de resultaten. Ten derde is de methodologie niet altijd even duidelijk omschreven. Zo laat Jansen bijvoorbeeld in het midden hoe hij de variabele *bepaaldheid van de VzC* geoperationaliseerd heeft, maakt hij geen onderscheid tussen zinsdelen en zinsdeelstukken en is het ook onduidelijk of hij inherente, niet-verplaatsbare VzC's meeneemt in zijn onderzoek⁴⁶. Ook de manier waarop *formaliteit* geoperationaliseerd wordt, is volgens mij betwistbaar en daarmee ook de verklaring die hij aan deze variabele verbindt. Bij een (wetenschappelijk) interview wordt de situatie niet minder of meer formeel door de vraagkeuze of door de mogelijkheid om langere antwoorden te formuleren (zeker met de beperkte tijdsduur van het interview, nl. 20 à 25 minuten, in het achterhoofd). Wat de situatie wel zou kunnen beïnvloeden is de relatie interviewer – geïnterviewde. In het artikel vermeldt Jansen zelf de interviews afgenomen te hebben bij 40 Leidenaars; zijn relatie met de geïnterviewden blijft echter onbesproken.

⁴⁵ “De sprekers kwamen uit vier groepen (elk van tien personen): mannen uit de twee hoogste sociale klassen (voortaan HK) (zie de *Beroepenklapper* van het ITS in Nijmegen), vrouwen uit diezelfde klassen; en mannen en vrouwen uit de laagste sociale klasse (voortaan LK)” (Jansen 1978: voetnoot 1).

⁴⁶ Jansen vermeldt wel dat bepaalde VzC's niet in extrapositie kunnen staan (1978: 76) maar hij voegt er niet aan toe of hij met deze verplaatsbaarheidsregel rekening houdt.

2.2.4.1.2 Braecke (1990)

Corpus: gesproken taalgebruik van 36 respondenten uit Opwijk (3 leeftijdscategorieën: 18-21 jaar, 35-45 jaar en 60+)

Methode: vragen over

- huidig of vroeger werk;
- vrijetijdsbesteding;
- indien oud genoeg: over de oorlog, het verschil tussen Opwijk in 1980 versus Opwijk vroeger en gevaarlijke situaties tijdens de oorlog

Net zoals Jansen (1978) onderzocht ook Braecke (1990) taalinterne en -externe variabelen. Hij maakt in tegenstelling tot Jansen geen onderscheid tussen formele en informele taal en zorgt ervoor dat “er niet formeel ‘ondervraagd’ werd” (Braecke 1982-1983: 114). In het onderzoek naar de taalinterne variabelen is de *focus*⁴⁷ volgens Braecke de alles verklarende variabele. In die mate zelfs dat andere variabelen ondergeschikt zijn aan *focus*. Braecke vergelijkt daarvoor gesproken data van 36 sprekers uit de Vlaams-Brabantse gemeente Opwijk en toont aan dat niet alleen voorzetselvoorwerpen maar ook andere bepalingen⁴⁸ vaker achteraan staan. Hoewel de verschillende bepalingen en de voorzetselvoorwerpen semantisch (en grammaticaal) zeer heterogeen zijn, hebben ze volgens Braecke (1990: 129) focusgeschiktheid als gemeenschappelijk kenmerk. Daarnaast merkt Braecke ook binnen de groep van bepalingen een verschil op. Hij stelt vast dat semantische klassen zoals *beperking*, *doel* en *middel* vaker achteraan staan dan bepalingen van *tijd* en *plaats*.

In tegenstelling tot Jansen (1978, 1979) stelt Braecke vast dat het aantal zinsdelen binnen de tang geen invloed heeft op de positie van VzC's. De *lengte van de VzC* beïnvloedt de positie wel, maar volgens Braecke is deze variabele ondergeschikt aan de veel prominentere variabele *focus*⁴⁹: “langere PPs staan meer in focus en langere focus PPs worden meer achteraan geplaatst” (Braecke 1990: 131). De positie van VzC's is dus volgens Braecke uitsluitend pragmatisch gemotiveerd en niet zozeer door de productieve of receptieve verwerking van de structuur. Deze redenering nuanceert Jansens idee, die in het effect van een pragmatische variabele én in het effect van de complexiteit gelooft.

⁴⁷ Een VzC krijgt het label *focus* als deze VzC “door de spreker een voorgrondrol wordt toegekend” (Braecke 1990: 129). Een nadeel volgens Braecke is dat deze variabele subjectief is, een voordeel is dat hij niet samenvalt met de onderverdeling ‘oud’ en ‘nieuw’.

⁴⁸ Een bijwoordelijke bepaling van richting, een tweede object, een handelend voorwerp, een bijwoordelijke bepaling van restrictie en reden, oorzaak en doel worden in Braeckes onderzoek allemaal onder de noemer ‘bepaling’ gerekend.

⁴⁹ In zijn proefschrift beschrijft Braecke zinsfocus als volgt: “Met focus duidt de spreker de relatieve prominentie aan van één deel in de uiting via het zinsaccent. Dat deel is een syntactische eenheid die refereert aan een element in de besproken stand van zaken. De keuze van dat element gebeurt autonoom door de spreker ten behoeve van de luisteraar. Bijgevolg kan de plaats van het zinsaccent nooit uit de syntactische structuur te voorspellen zijn.” (Braecke 1982-1983: 124).

Het moet echter nog vastgesteld worden of enkel deze discoursgemotiveerde variabele de positie van VzC's bepaalt dan wel of ook complexiteit een rol speelt en in welke mate de verschillende variabelen samen de woordvolgorde beïnvloeden.

Verder merkt Braecke op dat jongeren (de eerste leeftijdscategorie: 18- 21 jaar) meer extraponeren in vergelijking met de andere twee leeftijdscategorieën. Hij vergelijkt zijn resultaten ook met de resultaten van Jansen (1978) en stelt vast dat zijn informanten uit Opwijk (Braecke 1990) iets meer achterveldpositie gebruiken dan de respondenten uit Leiden (Jansen 1978). Toch blijft ook het middenveld in Opwijk de voorkeur genieten; de verhouding middenveld/achterveld in het onderzoek van Braecke is ongeveer 60/40, in dat van Jansen was dat 67,2%/32,8% voor de informele stijl). De reden waarom Vlamingen iets meer extraponeren, is volgens Braecke historisch te verklaren. De Zuid-Nederlandse syntaxis heeft de oudere SVO-volgorde bewaard (cf. de Meersman 1980). Braeckes resultaat kan echter ook aan de hand van de leeftijd verklaard worden. Braeckes respondenten werden namelijk in drie categorieën onderverdeeld (18-21, 35-45 en 60+), terwijl Jansens respondenten een leeftijd van 50 jaar of ouder delen. Bovendien stelt Braecke zelf vast dat jongeren meer extraponeren. Het verschil tussen Noord-Nederlands en Zuid-Nederlands lijkt er dus geen in termen van landsgrenzen te zijn, wel eentje op basis van leeftijd. Indien Braecke een verschil tussen Nederland (Leiden) en Vlaanderen (Opwijk) wil vaststellen, dan zou hij zijn derde leeftijdscategorie moeten vergelijken met de resultaten van Jansen.

Net zoals bij het onderzoek van Jansen kunnen we enkele methodologische beperkingen opsommen. Zo beperkt het onderzoek van Braecke zich tot een eerder kleine dataset (36 respondenten onderverdeeld in drie leeftijdscategorieën en opgesplitst volgens sekse; per leeftijdscategorie en per sekse 6 participanten). Ook Braecke zelf merkt deze beperking op: "Later contact met de specialisten in de sociale wetenschappen heeft ons van onze aanvankelijke euforie genezen, die ons liet denken dat taalgedrag zo een grote regelmaat vertoonde dat dit ook bij drie proefpersonen zichtbaar zou worden" (Braecke 1982-1983: 112). Er ontbreekt ook een multifactoriële statistische analyse, waardoor Braeckes bewering, namelijk dat *focus* de allesbepalende verklaring is, geen statistische ondersteuning krijgt. Het is ook in het onderzoek van Braecke onduidelijk welke types VzC's deel uitmaken van het onderzoeksobject en we kunnen de resultaten van het onderzoek in vraag stellen. Voor de operationalisatie van de discoursstatus⁵⁰ werkt Braecke met een combinatie van de grammaticale functie van een VzC en de semantische classificatie van een VzC. Hij stelt vast dat voornamelijk voorzetselvoorwerpen en bepalingen van *beperking*, *doel* en *middel* vaker achteraan staan.

⁵⁰ Braecke onderzocht net zoals Jansen ook het effect van *definietheid* (1. bepaalde VzC's, 2. onbepaalde VzC's en 3. VzC's met een bijwoord als een aparte categorie) en stelt vast dat definietheid wel een rol speelt bij vooraanplaatsing maar niet bij extrapositie.

Volgens Braecke zijn deze VzC's vooral VzC's met focusgeschiktheid en staan ze daarom achteraan. Dat roept tegelijkertijd de volgende vraag op: hebben voorzetselvoorwerpen en bepalingen van bepaalde semantische klassen (zoals *beperking*, *reden*, *oorzaak*, enz.) meer focusgeschiktheid en staan ze daarom meer in het achterveld of staan ze meer in het achterveld en krijgen ze daarom het label *focusgeschikt*? De operationalisatie van *focus* (of van *focusgeschikt*) kan dus heel subjectief zijn, wat Braecke trouwens zelf toegeeft (Braecke 1990: 129), waardoor we nog altijd in het ongewisse blijven op welke manier deze discoursgemotiveerde variabele een rol speelt. In huidig onderzoek willen we de operationalisatie van de discoursstatus uitbreiden door ook rekening te houden met de rest van het discours en door de voorspelbaarheid en/of toegankelijkheid van de VzC in kaart brengen (cf. Grondelaers & Speelman 2007; Grondelaers et al. 2009).

2.2.4.2 Psycholinguïstisch geïnspireerd onderzoek: Jansen (2003)

Corpus: schrijftaalcorpus op basis van 5 journalistieke teksten:

- 2 voetbalberichten en 3 columns
- in totaal 181 zinnen

In Jansen (2003) gaat de auteur na of de *sentence-processing* theorieën van Hawkins (1994) en Gibson (1998, 2000) de positie van complexe constituenten voorspellen. Complex wordt in deze studie een constituent genoemd die een relatieve bijzin bevat of minstens een VzC als nabepaling heeft. We merken hierbij onmiddellijk op dat het niet duidelijk is hoeveel (complexe) VzC's onderzocht werden, uit welke krant de artikelen komen en of de artikelen door een of meerdere auteurs geschreven zijn.

Jansen voorspelt puur intuïtief dat een complexe constituent, zoals een complexe VzC's, het middenveld vermijdt en een positie buiten de tang verkiest. Stilistisch gezien is volgens Jansen de positie na de tang (of achterveldpositie) te verkiezen boven de positie voor de tang (of voorveldpositie). Om deze intuïtie te staven, telt Jansen het aantal eenvoudige⁵¹ en complexe constituenten in de drie mogelijke posities (voorveld, middenveld en achterveld). De resultaten tonen dat complexe constituenten het middenveld vermijden in vergelijking met eenvoudige constituenten (22% van de complexe staan in het middenveld versus 65% van de eenvoudige) en dat complexe constituenten het achterveld boven het voorveld verkiezen (69% in achterveld versus 31% in voorveld). Jansen vertrekt dus vanuit een (voorkeurspositie)hiërarchie: eerst achterveldpositie, dan voorveldpositie, en tot slot, als de andere posities niet ingenomen kunnen worden, middenveldpositie. Jansen gebruikt Hawkins *Principle of Early Immediate Constituents* (1994) en Gibsons *Syntactic Prediction Locality Theory* (1998) om deze volgorde psycholinguïstisch te motiveren.

⁵¹ Eenvoudige constituenten hebben noch een betrekkelijke bijzin als nabepaling, noch een VzC als nabepaling.

Hawkins beschouwt woordvolgorde als een manier om de syntactische structuur van een zin zo snel en efficiënt mogelijk te analyseren. Om een volledig overzicht te krijgen van de zinsstructuur hoeft de toehoorder niet te wachten tot het laatste woord van de zin. Zodra de toehoorder genoeg informatie heeft over hoe de laatste constituent is opgebouwd, kan de structuur van die laatste constituent en de hele zin bepaald worden. Volgens het *Principle of Early Immediate Constituents* (1994) krijgt die woordvolgorde de voorkeur die de toehoorder in staat stelt de zinsstructuur snel en efficiënt te bepalen. Dit principe kan geïllustreerd worden aan de hand van de voorbeeldzinnen uit Jansen (2003: 83):

- (45) a In de herfst van zijn loopbaan, waarin hij het vizier meer dan ooit gericht heeft op die ene medaille die hem nog ontbreekt, weet Rintje Ritsma zich voor een delicaat olympisch vraagstuk geplaatst. → **voorveld**
 b Rintje Ritsma weet zich voor een delicaat olympisch vraagstuk geplaatst in de herfst van zijn loopbaan, waarin hij het vizier meer dan ooit gericht heeft op die ene medaille die hem nog ontbreekt. → **achterveld**
 c Rintje Ritsma weet zich in de herfst van zijn loopbaan, waarin hij het vizier meer dan ooit gericht heeft op die ene medaille die hem nog ontbreekt, voor een delicaat olympisch vraagstuk geplaatst. → **middenveld**

In zin 45a en 45c moet de taalgebruiker wachten tot het laatste woord van de zin (*geplaatst*) om het werkwoord in combinatie met de argumenten te verwerken. In zin 45b komt het werkwoord veel sneller en weet de taalgebruiker vanaf het woord *in* dat er enkel nog een bijwoordelijke bepaling kan volgen. Daardoor is de woordvolgorde in 45b makkelijker te verwerken dan die in 45a en 45c. Hawkins theorie kan ook mathematisch berekend worden. De teller in onderstaande breuken stelt het aantal zinsdelen voor (6 zinsdelen in totaal), de noemer het aantal woorden (33 in totaal). Ter illustratie: in zin 45a is het vervoegde werkwoord *weet* het tweede zinsdeel en het 24^{ste} woord. Vervolgens worden deze breuken opgeteld en wordt de uitkomst van deze som gedeeld door het aantal breuken (of aantal zinsdelen). De woordvolgordevariant met de hoogste waarde is de beste variant.

- (45) a //In de herfst van zijn loopbaan, waarin hij het vizier meer dan ooit gericht heeft op die ene medaille die hem nog ontbreekt, $\frac{1}{23}$ // weet $\frac{2}{24}$ // Rintje Ritsma $\frac{3}{26}$ // zich $\frac{4}{27}$ // voor een delicaat olympisch vraagstuk $\frac{5}{32}$ // geplaatst $\frac{6}{33}$ //
 Som: $(1/23) + (2/24) + (3/26) + (4/27) + (5/32) + (6/33) = 0.73$
 $0.73/6$ (aantal zinsdelen) = **0.12**
 b //Rintje Ritsma $\frac{1}{2}$ // weet $\frac{2}{3}$ // zich $\frac{3}{4}$ // voor een delicaat olympisch vraagstuk $\frac{4}{9}$ // geplaatst $\frac{5}{10}$ // in de herfst van zijn loopbaan, waarin hij het vizier meer dan ooit gericht heeft op die ene medaille die hem nog ontbreekt $\frac{6}{33}$ //.
 Som: $(1/2) + (2/3) + (3/4) + (4/9) + (5/10) + (6/33) = 3.04$
 $3.04/6$ = **0.51**

c //Rintje Ritsma $\frac{1}{2}$ // weet $\frac{2}{3}$ // zich $\frac{3}{4}$ // in de herfst van zijn loopbaan, waarin hij het vizier meer dan ooit gericht heeft op die ene medaille die hem nog ontbreekt, $\frac{4}{27}$ // voor een delicaat olympisch vraagstuk $\frac{5}{32}$ // geplaatst $\frac{6}{33}$ //
 Som: $(1/2) + (2/3) + (3/4) + (4/27) + (5/32) + (6/33) = 2.40$
 $2.40/6 = 0.40$

Op basis van Hawkins procedure krijgt achterveldpositie de absolute voorkeur en is middenveldpositie makkelijker te verwerken dan voorveldpositie. Hawkins procedure zou de volgende hiërarchie veronderstellen: achterveld > middenveld > voorveld, wat niet helemaal overeenkomt met Jansens opvatting van achterveld > voorveld > middenveld.

Gibsons *Syntactic Prediction Locality Theory* (1998) is complexer dan Hawkins' theorie. Volgens Gibson gebeuren er twee activiteiten die elk een zekere verwerkingskost met zich meebrengen, namelijk een integratiekost: “the integration of every next word in the sentence into the structure built up so far” en een geheugenkost: “the storage in working memory of all elements that are needed to complement the elements processed so far for a meaningful grammatical sentence” (Jansen 2003: 87). Jansen (2003) berekent enkel de integratiekost.

De integratiekost hangt af van de afstand tussen een nieuw element en het hoofd(werkwoord). Elke nieuwe referent kost 1 *Integration Unit* (of IU). Hoe meer IUs er nodig zijn om een bepaalde woordvolgorde te verwerken, hoe moeilijker de woordvolgorde. De volgende voorbeeldzinnen (uit Jansen 2003: 88) illustreren Gibsons theorie:

1	In	de	herfst	van	zijn	Loopbaan	Ziet	Rintje	een	dilemma
	0	0	0	0	0	0	Hoofd	IU(1)	0	IU(2)
2	Rintje	ziet	een	dilemma	in	De	Herfst	van	zijn	loopbaan
	IU(1)	Hoofd	0	IU(1)	0	0	0	0	0	0
3	Rintje	ziet	in	de	herfst	Van	Zijn	loopbaan	een	dilemma
	IU(1)	Hoofd	0	0	0	0	0	0	0	IU(3)

Tabel 1: Voorstelling Gibsons theorie (uit Jansen 2003: 88)

In de eerste zin in tabel 1 weet de taalgebruiker dat het werkwoord *zien* een subject en een direct object oproept. Het volgende woord *Rintje*, het subject, is verbonden met het werkwoord. Om *Rintje* te verwerken, kost dat de taalgebruiker 1 IU. Vervolgens verwerkt de taalgebruiker de constituent *een dilemma*. *Een dilemma* is het direct object, eveneens verbonden aan het werkwoord *zien*. Deze keer is de integratiekost 2 IUs want de taalgebruiker moet voorbij de referent *Rintje*, vooraleer hij de referent *dilemma* bereikt. In de tweede zin staat *Rintje* voor het werkwoord. Wanneer de taalgebruiker het werkwoord leest of hoort, kan hij *Rintje* aan het werkwoord koppelen. Dat kost opnieuw 1 IU. Omdat *een dilemma* direct na het werkwoord staat, kost de integratie van het direct object ook 1 IU want de taalgebruiker moet niet voorbij een andere referent. In de derde zin kost de integratie van *Rintje* opnieuw 1 IU (zoals in zin 2) maar de integratie van een

dilemma kost 3 IUs. De reden is dat de taalgebruiker voorbij de referenten *herfst* en *loopbaan* moet vooraleer hij de referent *dilemma* bereikt.

De woordvolgorde met de minste IUs is zin 2, gevolgd door zin 1 en zin 3. De hiërarchie komt dus overeen met Jansens verwachting: achterveld > voorveld > middenveld. Alleen werkt deze berekening niet meer als er een tweede werkwoordvorm toegevoegd wordt en het hoofdwerkwoord de IUs bepaalt. In dat geval genieten zowel achterveldpositie (zin 4 in tabel 2) als voorveldpositie (zin 3 in tabel 2) de voorkeur boven middenveldpositie (zin 5 in tabel 2):

3	In	de	herfst	van	zijn	loopbaan	moet	Rintje	Een	dilemma	oplossen
	0	0	0	0	0	0	IU(3)	IU(2)	0	IU(1)	hoofd
4	Rintje	moet	een	dilemma	oplossen	in	De	herfst	Van	zijn	loopbaan
	IU(3)	IU(2)	0	IU(1)		0	0	0	0	0	0
5	Rintje	moet	in	de	herfst	van	zijn	Loopbaan	Een	Dilemma	oplossen
	IU(5)	IU(4)	0	0	0	0	0	0	0	IU(1)	

Tabel 2: Voorstelling Gibsons theorie bij een zin met twee ingevulde polen

Beide theorieën verklaren dus maar gedeeltelijk de voorkeurspositie van complexe constituenten. Hawkins' theorie voorspelt de voorkeur voor achterveldpositie, maar plaatst middenveldpositie voor voorveldpositie. Gibsons theorie werkt in een zin met één ingevulde pool. In een zin met twee polen, maakt Gibsons theorie geen onderscheid meer tussen voorveldpositie en achterveldpositie. Beide posities zijn wel te verkiezen boven middenveldpositie. Volgens Jansen (2003) ligt de oplossing in de manier waarop Gibsons integratiekost berekend wordt. In de oorspronkelijke berekening hangt de integratiekost af van het hoofdwerkwoord (het werkwoord *ziet* in tabel 1 en *oplossen* in tabel 2). Een Nederlandse taalgebruiker hoeft echter niet altijd tot het hoofdwerkwoord te wachten vooraleer hij de zin kan begrijpen. Vooral de integratie van het subject gebeurt volgens Jansen (2003) anders dan in Gibsons voorstel. Wanneer de taalgebruiker namelijk beseft dat de eerste constituent niet het subject is, kan hij veronderstellen dat de constituent na het vervoegde werkwoord het subject zal zijn. De integratie van het subject hoeft dus niet te wachten tot het hoofdwerkwoord *oplossen*, maar kan eigenlijk na het vervoegde werkwoord *moet* plaatsvinden. Op basis van de nieuwe berekening krijgt extrapositie ook in een zin met twee ingevulde polen de absolute voorkeur:

3	In	de	herfst	van	zijn	loopbaan	moet	Rintje	Een	dilemma	oplossen
	0	0	0	0	0	0	IU(3)	IU(1)	0	IU(1)	
4	Rintje	moet	een	dilemma	oplossen	In	de	herfst	Van	zijn	loopbaan
	IU(1)	IU(2)	0	IU(1)		0	0	0	0	0	0
5	Rintje	moet	in	de	herfst	van	zijn	Loopbaan	Een	Dilemma	oplossen
	IU(1)	IU(4)	0	0	0	0	0	0	0	IU(1)	

Tabel 3: Voorstelling Gibsons theorie met aanpassing van Jansen (2003: 90)

Gibsons *Syntactic Prediction Locality Theory* toont bijgevolg aan wat Jansen intuïtief veronderstelt en empirisch aantoonde: complexe constituenten, zoals een complexe VzC, staan in eerste instantie in het achterveld, op de tweede plaats in het voorveld en pas als laatste optie in het middenveld.

2.2.4.3 Corpusgebaseerd onderzoek naar de positie van voorzetselvoorwerpen: Hoeksema (2014)

Corpus: schrijftaalcorpus bestaande uit gedrukte en online bronnen (www.delpher.nl en www.dnbl.org)

- uitsluitend prozateksten (deels fictie en deels non-fictie)
- in totaal 3400 zinnen uit de periode 1700-2014 (2269 hoofdzinnen, 1022 bijzinnen en 109 ja/neen-vragen)

Hoeksema (2014) bestudeert de variabele positie van voorzetselvoorwerpen (voorzetselobjecten of VzO's) in een schrijftaalcorpus. Hij beperkt zich echter niet tot middenveldpositie versus achterveldpositie maar houdt rekening met 4 mogelijke posities:

- (46)
- a Topicalisatie: **Aan mij** zal het niet liggen.
 - b Scrambling: Het zal **aan mij** niet liggen.
 - c Regulier: Het zal niet **aan mij** liggen.
 - d Extrapositie: Het zal niet liggen **aan mij**.

In het middenveld kan het voorzetselvoorwerp dus links (scrambling) of rechts (regulier) verschijnen. Hoeksema veronderstelt daarbij dat “het voorzetselvoorwerp als complement van het werkwoord in de VP staat, en door scrambling over bepalingen heen getild kan worden die normaliter ter linkerzijde van het voorwerp zouden staan” (Hoeksema 2014: 221; zie ook Broekhuis & Corver 2016: 1551).

Hoeksema voert enerzijds een diachrone studie uit naar de plaatsing van voorzetselvoorwerpen, zie ook paragraaf 2.2.1.2, en is anderzijds ook geïnteresseerd in de synchrone aspecten van deze variatie. Voor dat laatste beperkt hij zijn corpusmateriaal tot recenter geschreven proza van 1950-2014 en bestudeert hij de invloed van *de lengte van de VzC in lettergrepen, de definietheid van het voorzetselvoorwerp en de lexicale eenheid van werkwoord en voorzetsel*.

Op basis van zijn onderzoek stelt Hoeksema vast dat extrapositie een van de minst gebruikte opties is bij korte voorzetselvoorwerpen (tot en met 3 lettergrepen). Vooral de reguliere positie geniet dan de voorkeur (in 62,5% van de onderzochte zinnen staat een kort voorzetselvoorwerp in een reguliere positie, slechts in 8% in extrapositie). Bij middellange voorzetselvoorwerpen (4 tot en met 8 lettergrepen) gaat extrapositie ongeveer gelijk op met de reguliere positie (112 in extrapositie of 37,6% en 115 in reguliere

positie of 38,6%). Extrapositie geniet duidelijk de voorkeur bij lange voorzetselvoorwerpen, bestaande uit 9 lettergrepen of meer (in 68,6% van de zinnen).

Voor de definietheid maakt Hoeksema een onderscheid tussen “persoonlijke voornaamwoorden, R-pronomina, definiete NPs met determinator, eigennamen, indefiniete NPs met een determinator en kale nomina” (Hoeksema 2014: 230), waarbij hij dus een hiërarchie creëert van zeer definiet tot zeer indefiniet. Wat de extrapositie van deze voorzetselvoorwerpen betreft, merkt hij op dat persoonlijke voornaamwoorden en R-pronomina nooit in extrapositie staan en dat voornamelijk definiete NPs en indefiniete NPs met een determinator achterveldpositie verkiezen (resp. 53% en 57%), meer zelfs dan kale nomina (41%). De resultaten zijn volgens Hoeksema gerelateerd aan de gemiddelde lengte van definiete NPs, indefiniete NPs en kale nomina: definiete zijn gemiddeld 8,6 lettergrepen lang, indefiniete gemiddeld 9,2 lettergrepen en kale nomina ‘slechts’ 6,7 lettergrepen. In een bijkomende proef selecteert Hoeksema enkel zinnen met definiete voorzetselvoorwerpen met minimaal 3 en maximaal 5 lettergrepen. De gemiddelde lengte is dan 4,1 lettergrepen en het aantal extraposities daalt sterk tot ongeveer 27%.

Tot slot bestudeert Hoeksema de lexicale verschillen tussen *denken aan* en *houden van*. Hij vertrekt vanuit het idee dat lexicale combinaties van een werkwoord en voorzetsel verschillen laten zien in de volgordevoorkeuren van voorzetselvoorwerpen (Hoeksema 2014: 222). Beide combinaties tonen een voorkeur voor de positie *regulier/extrapositie*. Dit label verwijst naar zinnen waarbij de tweede pool niet ingevuld is, waardoor het niet duidelijk is of een voorzetselvoorwerp deel uitmaakt van het (einde van het) middenveld, dan wel van het achterveld. Met deze positie kunnen we bijgevolg geen rekening houden. We zien wel een grotere voorkeur voor extrapositie bij de combinatie *denken aan* (17%) dan bij de combinatie *houden van* (6%), maar of daardoor de lexicale combinatie een directe rol speelt bij de plaatsing van voorzetselvoorwerpen blijft nog een open vraag.

We kunnen besluiten dat het synchrone onderzoek van Hoeksema aangetoond heeft dat vooral de lengte de positie van voorzetselvoorwerpen bepaalt. Zelfs in die mate dat ‘afwijkende’ resultaten bij de variabele definietheid door de complexiteit verklaard kunnen worden.

2.2.4.4 Conclusie empirische studies

In onderstaande tabel geven we een opsomming van de vermelde studies (kolommen) en de onderzochte variabelen (rijen). Met behulp van deze tabel krijgen we een duidelijk overzicht welke variabelen reeds onderzocht werden:

	Jansen (1978, 1979)	Braecke (1990)	Hoeksema (2014)
grammaticale functie VzC	X	X	
accentuering VzC	X		
definietheid VzC	X		X (enkel VzO)
complexiteit VzC	X	X	X (enkel VzO)
lengte MV	X	X	
zinstype	X		
aanwezigheid anafoor	X		
stijl	X		
seks	X		
klasse	X		
focus		X	
leeftijd		X	

Tabel 4: Overzicht van de onderzochte variabelen in empirisch onderzoek (§2.2.4)

In een volgende tabel geven we een samenvatting van de variabelen die volgens de studies een duidelijke voorkeur hadden voor een van de twee plaatsingsvoorkeuren:

	Middenveldpositie	Achternveldpositie
grammaticale functie VzC		voorzetselvoorwerp + bepalingen van <i>beperking</i> , enz.
definietheid VzC	definiet	Indefiniet
complexiteit VzC		complexere of langere VzC's
lengte MV		langere MV
stijl	informeel	Formeel
seks	vrouwen	Mannen
klasse	hogere klasse	lagere klasse
leeftijd	oud	Jong

Tabel 5: Overzicht variabelen op basis van het bestaande empirische onderzoek (§2.2.4) met hun voorkeur voor middenveld- of achternveldpositie

Samengevat wil elke empirische studie die we in paragraaf 2.2.4 besproken hebben, vat krijgen op de mechanismen die de keuze voor een van de volgordevarianten bepalen. Door de invloed van allerlei taalinterne en -externe variabelen te bestuderen en op zoek te gaan naar psycholinguïstische motivaties, heeft men vastgesteld dat de keuze tussen middenveldpositie en achternveldpositie beïnvloed wordt door meerdere (eventueel zelfs interagerende) variabelen. Ondanks het intensieve onderzoek is het echter nog steeds onduidelijk welke variabelen prominenter zijn dan andere, welke variabelen eventueel interageren en wat het functionele onderscheid is tussen deze twee volgordes. Daar liggen meerdere redenen aan ten grondslag.

Vooreerst is het gebruikte taalmateriaal in de werken van Jansen (1978, 1979) en Braecke (1990) onvoldoende afgebakend en/of niet duidelijk gepresenteerd, waardoor een onderlinge vergelijking en replicatieonderzoek moeilijk is. Het psycholinguïstische onderzoek van Jansen (2003) toont aan dat complexe constituenten het achterveld verkiezen boven het voorveld en het middenveld, maar het is niet duidelijk of enkel complexe VzC's onderzocht werden, dan wel alle complexe constituenten (zoals een complexe NC, AdjC, enz.). Het lijkt nochtans evident dat het constituententype en de grammaticale functie eveneens bepalen welke constituenten in achterveldpositie kunnen staan. Voorts maakt enkel Hoeksema een onderscheid tussen zinsdelen en zinsdeelstukken (voorzetselvoorwerpen zijn onafhankelijke VzC's), terwijl de afhankelijkheid ook de positie in de zin mee kan beïnvloeden.

Naast deze methodologische opmerkingen kunnen we uit het pioniersonderzoek van Jansen (1978, 1979) en Braecke (1990) opmerken dat het vooralsnog onduidelijk is welke rol *syntactische complexiteit* en *discoursstatus* spelen. In navolging van Jansen (1978) zouden we kunnen beweren dat zowel de *complexiteit* als de *discoursstatus* bepalen of een VzC in middenveldpositie of in achterveldpositie staat. Volgen we Braecke, dan speelt enkel de *discoursstatus* een rol en is *complexiteit* hieraan ondergeschikt. De oorzaak van deze onduidelijkheid is het gebrek aan statistische toetsing. Beide onderzoeken zijn namelijk monofactorieel waardoor enkel het effect van een variabele op de positie van de VzC bestudeerd wordt. Het is dus nog een open vraag hoe de verschillende variabelen zich tot elkaar verhouden in hun effect op de plaatsing en of deze variabelen de volledige variatie in plaatsingsgedrag kunnen verklaren. Het gebrek aan statistische toetsing staat bijgevolg een algemene verklaring in de weg. Het onderzoek van Hoeksema 2014 wijst ook in een andere richting en stelt dat de *lexicale combinatie van een werkwoord en een vast voorzetsel* een plaatsingsvoorkeur kan oproepen.

Deze studie wil samengevat achterhalen op welke manier de onderzochte en mogelijks verklarende variabelen een rol spelen, of *complexiteit* en *discoursstatus* interageren, of het *werkwoord* al dan niet de keuze tussen middenveldpositie en achterveldpositie mee bepaalt en hoeveel variatie deze variabelen kunnen voorspellen. Daarvoor zullen we een multivariate statistische techniek gebruiken, m.n. een logistische-regressieanalyse.

Hoofdstuk 3

Methodologie

The issue of operationalization is one of the most important of all. If you do not operationalize your variables properly, then the whole study might be useless since you may actually end up not measuring what you want to measure. (Gries 2013: 18)

In dit hoofdstuk worden de methodologische grondslagen van ons onderzoek naar de variabele plaatsing van VzC's besproken. Paragraaf 3.1 licht eerst de inhoud en structuur van het Dutch Parallel Corpus toe (3.1.1) en vervolgens de selectiecriteria die ons in staat zullen stellen om op een systematische, weloverwogen manier de relevante VzC's uit het corpus te halen (3.1.2). Hoewel we dit onderzoek uit praktische overwegingen beperken tot slechts één tekstsoort (m.n. de journalistieke) en tot Belgisch-Nederlands, testen we of de subgenres binnen de journalistieke teksten en het verschil tussen vertaald en niet-vertaald Nederlands (een essentieel aspect van het Dutch Parallel Corpus) een effect hebben op de variabele positie van VzC's. De effecten van het subgenre en van (niet)-vertaald Nederlands worden respectievelijk in paragraaf 3.1.3 en 3.1.4 besproken.

In paragraaf 3.2 gaat de aandacht naar de onderzochte variabelen en de manuele validatie van deze variabelen. De opname van precies deze variabelen is uiteraard niet willekeurig gebeurd. We baseren ons hiervoor op de reeds besproken literatuur in hoofdstuk 2 en op soortgelijk onderzoek naar andere types van woordvolgordevariatie in het Nederlands en het Engels. Op basis van deze literatuur kunnen we ook al enkele hypothesen formuleren over het effect van deze variabelen op de positie van VzC's in het Nederlands.

3.1 Corpus en dataselectie

3.1.1 Dutch Parallel Corpus

Het Dutch Parallel Corpus (afgekort DPC, Macken et al. 2011) bevat 10 miljoen woorden voor de talenparen Nederlands-Engels en Nederlands-Frans, met het Nederlands als centrale taal. Dit bidirectionele parallelle corpus trekt door de aanwezigheid van vertaalde en niet-vertaalde taal vooral de aandacht van vertaalkundige studies (o.a. Delaere 2015 en Vandevorde 2016), maar is ook voor onderzoek binnen de neerlandistiek zonder een vertaalkundig perspectief waardevol. Het DPC bevat namelijk een grote hoeveelheid niet-vertaald (Belgisch- en Nederlands-) Nederlands, verdeeld over zes tekstsoorten (*administratieve teksten*, *instructieve teksten*, *journalistieke teksten*, *externe communicatie*, *non-fictie* en *fictie*) en nodigt zo uit om een taalfenomeen in één of meerdere tekstsoorten te bestuderen. Het is ook zinsgealigneerd en gelemmatiseerd.

Voor onderhavig onderzoek selecteren we één tekstsoort, namelijk journalistieke teksten en één taalvariant, namelijk Belgisch-Nederlands. In tabel 6 geven we een overzicht van het aantal tokens en het aantal teksten voor de Belgisch-Nederlandse journalistieke component, het aantal subgenres binnen de journalistieke teksten en of ze vertaald en niet-vertaald Nederlands bevatten (uit Delaere 2015: 45-46 en Delaere 2015: 53).

Journalistieke teksten		Origineel Nederlands		Vertaald Nederlands uit Frans		Vertaald Nederlands uit Engels	
		485.876 tokens	356 teksten	272.429 tokens	240 teksten	295.039 tokens	253 teksten
Subgenres	Opiniestukken	✓		✓		✓	
	Nieuwsartikelen	✓				✓	
	Informatieve documenten	✓					

Tabel 6: Overzicht journalistieke teksten in het DPC (uit Delaere 2015)

Onderzoek van Delaere (2015) heeft echter aangetoond dat niet alle data binnen de tekstsoort *journalistieke teksten* in het Dutch Parallel Corpus ook effectief journalistieke teksten zijn. Zo vallen heel wat teksten onder de tekstsoorten *instructieve teksten* en *non-fictie*. We hebben daarom rekening gehouden met Delaeres reorganisatie van het DPC op basis van de afzender, de geadresseerde, het gebruik en het communicatieve doel.

Er zijn meerdere redenen waarom we met journalistieke teksten werken. Ten eerste zijn journalistieke teksten gemakkelijk te verzamelen, waardoor vervolgonderzoek mogelijk is. Ten tweede hebben journalistieke teksten eenzelfde, specifieke doel voor ogen, namelijk de lezer informeren en opiniëren over de actualiteit. Daarvoor zullen de

schrijvers een actuele taal hanteren die door een breed publiek begrepen kan worden. De tekstsoort levert bijgevolg data op die het professionele hedendaags geschreven taalgebruik weerspiegelt (een tegenvoorbeeld is administratieve taal die vaak als zeer archaïsch, zowel lexicaal als grammaticaal, wordt bestempeld). Journalistieke taal en literaire taal worden bovendien vaak gebruikt in studies naar syntactische variatie. De Sutter & Van de Velde (2008), bijvoorbeeld, onderzoeken de positie van VzC's in vertaalde en niet-vertaalde literaire taal (Nederlands en Duits). Door in deze studie journalistieke taal te gebruiken, kunnen we een ander tekstgenre belichten. Tot slot kunnen we door te focussen op één genre, een eventueel genre-effect uitsluiten. Het tekstgenre op zich kan namelijk de positie van VzC's beïnvloeden. Zo vermelden stijlgidsen dat vooral juridische teksten en overheidsteksten middenveldpositie verkiezen (o.a. Zuidgeest 1995; zie §2.1.2). Door slechts één tekstgenre te selecteren, kan de positie van VzC's al niet het gevolg zijn van het overkoepelende tekstgenre an sich (wel kan er binnen het tekstgenre *journalistieke teksten* nog een verschil optreden tussen de subgenres, zie daarvoor §3.1.3). Een beperking is wel dat huidig onderzoek enkel iets meer kan zeggen over de woordvolgorde van VzC's in Belgisch-Nederlandse journalistieke geschreven teksten.

De eerste reden waarom we enkel met Belgisch-Nederlands werken, ligt in lijn met de derde motivatie om slechts met één tekstsoort te werken. Net zoals de verschillende tekstsoorten een effect kunnen hebben op de positie van VzC's, kan de positie van VzC's ook in de Nederlandse variëteiten verschillen. Braecke (1990) bevestigt deze hypothese; hij merkt een verschil op in de positie van VzC's in het Nederlands in Leiden (op basis van het onderzoek van Jansen 1978) en in het Nederlands in Opwijk. De tweede, en meest doorslaggevende reden is dat het aandeel Nederlands-Nederlands in het DPC beperkt is. Bijgevolg kunnen we deze component onvoldoende bestuderen in vergelijking met de Belgisch-Nederlandse variant.

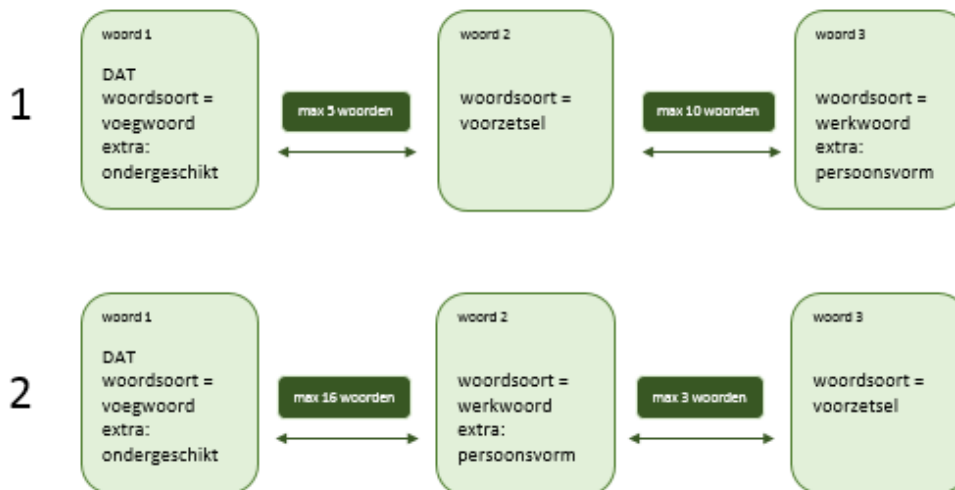
3.1.2 Selectiecriteria

In de volgende subparagrafen bespreken we de vijf selectiecriteria waarmee we rekening hebben gehouden. Nadat alle selectiecriteria werden toegepast, hielden we een dataset van 1725 zinnen over.

3.1.2.1 De afstand tussen het bindwoord en het vervoegde werkwoord is maximaal 16 woorden

In dit onderzoek zijn we enkel geïnteresseerd in bijzinnen omdat de tweede pool (of werkwoordelijke eindgroep) in dat geval zichtbaar is en we een duidelijke scheiding krijgen tussen middenveld en achterveld. Om zoveel mogelijk relevante bijzinnen te selecteren met een VzC in een van beide posities en om de hoeveelheid data enigszins te

beperken, limiteren we de afstand tussen 1. het bindwoord¹ en een voorzetsel (om zoveel mogelijk zinnen met een VzC in het middenveld te selecteren) en tussen 2. het vervoegde werkwoord en een voorzetsel (om zoveel mogelijk zinnen met een VzC in het achterveld te selecteren).² De zoekopdrachten geven we schematisch weer in figuur 6:



Figuur 6: Schematisch voorstelling van het eerste selectiecriteria: *de afstand tussen het bindwoord en het vervoegde werkwoord is maximaal 16 woorden*

In het eerste geval verwachten we na maximaal 5 woorden³ na het bindwoord een voorzetsel (tag: VZ(init)⁴). De afstand tussen het voorzetsel en het vervoegde werkwoord (tag: WW(persoonsvorm)) laten we ook variëren en kan tot maximaal 10 woorden oplopen. De maximale afstand tussen het bindwoord en het vervoegde werkwoord is dus 16 woorden lang, wat op zich een zware tangconstructie is.

Om een VzC in het achterveld te selecteren, verwachten we maximaal 3 woorden na het vervoegde werkwoord een voorzetsel. Volgens de ANS is de aanwezigheid van meer dan één zinsdeel (of zinsdeelstuk) in het achterveld niet gebruikelijk en staat een geëxtraponeerde VzC meestal voor een geëxtraponeerde bijzin. Een maximale afstand van 3 woorden is dus voldoende om een relevant voorzetsel te selecteren. Ook in deze

¹ Zie ook selectiecriteria 2, §3.1.2.2

² In het geval we enkel zoeken naar zinnen waarin het voegwoord *dat* binnen een afstand van max. 16 woorden gevolgd wordt door een vervoegd werkwoord, krijgen we 23689 attestaties. Deze attestaties zouden vervolgens manueel gecontroleerd moeten worden op de aanwezigheid van een VzC in combinatie met de selectiecriteria die in de volgende paragrafen worden besproken. Om de dataset enigszins te beperken, hebben we de zoekopdracht aangepast.

³ Nadat we de zinnen geanalyseerd hadden, hebben we de tang ter illustratie en voor de zekerheid vergroot tot maximaal 10 woorden. Het resultaat waren 157 nieuwe zinnen waarvan we de eerste 100 gecontroleerd hebben. Van deze 100 zinnen waren slechts 13 zinnen relevant. Een grotere tang zou dus niet automatisch tot een significant hoger aantal zinnen met een VzC in het middenveld hebben geleid.

⁴ *Init* staat voor *initieel*.

zoekopdracht is de maximale afstand tussen het bindwoord en het vervoegde werkwoord 16 woorden (zie ook zoekopdracht 2 figuur 8).

Uiteraard zorgen de zoekopdrachten naar VzC's in het middenveld en in het achterveld voor herhalingen die in een volgende stap geëlimineerd worden. Alle zinnen mogen tenslotte slechts één keer in de dataset voorkomen. De dataset bestaat uit 5135 zinnen die aan verdere selectiecriteria werden onderworpen en manueel werden gecontroleerd. Door bijkomende selectiecriteria in te voeren, vermijden we dat we met een te complexe dataset werken waarbij de variabelen die in deze studie aan bod komen vertekend zouden worden door andere, verborgen covariërende variabelen waar we voorlopig niet in geïnteresseerd zijn. Ter illustratie: in onderstaande voorbeeldzinnen staat de VzC telkens in een bijzin, toch worden ze niet meegenomen in het onderzoek. De reden daarvoor is dat de VzC geen deel uitmaakt van een grammaticale bijzin ingeleid door een grammaticaal onderschikkend voegwoord *dat* (zin 47), de VzC geen zinsdeel is (maar een zinsdeelstuk; zin 48), de VzC inherent verbonden is met het werkwoord (en daardoor niet in het achterveld kan staan volgens het inherentieprincipe (ANS 1997); zin 49), de VzC voorafgegaan wordt door een focuspartikel (zin 50), het een voornaamwoordelijk bijwoord is (zin 51), een beknopte bijzin (zin 52) of deel uitmaakt van een *aan het + infinitief*-constructie (zin 53).

- (47) Een reclamebureau in Brussel **dat** onze internetdiensten zou willen integreren in zijn aanbod voor klanten
- (48) Verschillende internationale onderzoeken suggereren dat het IQ **van de bevolking** de laatste jaren is toegenomen.
- (49) We hebben alleen willen aantonen dat het Italiaanse voetbal **van een zeer hoogstaand niveau** is.
- (50) Uit deze VRWB-studie blijkt verder dat een promotie **vooral in de natuur- en toegepaste wetenschappen en de medische wetenschappen** een maatschappelijke meerwaarde heeft, terwijl dit voor de cultuur- en gedragswetenschappen nauwelijks het geval is.
- (51) Praat over onderwerpen waarvan u vermoedt dat uw gesprekspartner **er** geïnteresseerd **in** is.
- (52) ... dat zij China veel schade zouden kunnen berokkenen **door simpelweg het 'cn'-domein weg te halen**.
- (53) Tal van enquêtes wijzen erop dat de voorraadafbouw **aan het slinken** is ...

Strikt afgebakende selectiecriteria zorgen ervoor dat de repliceerbaarheid en continuïteit in toekomstig onderzoek gegarandeerd wordt, wat tevens een methodologische doelstelling van dit onderzoek is.

3.1.2.2 De VzC komt voor in een bijzin ingeleid door een grammaticaal onderschikkend voegwoord *dat*

We beperken het type bijzin in ons onderzoek tot grammaticale voegwoorzinnen ingeleid door het grammaticaal onderschikkend voegwoord *dat*, omdat de bijzin zelf een rol kan spelen en een effect kan hebben op de variabele positie van VzC's. Door één type bijzin te selecteren, sluiten we een eventueel bijzinstype-effect uit.

We gebruiken de woordsoortinformatie (of tag) die in het Dutch Parallel Corpus opgenomen is. Om zinnen met een onderschikkend voegwoord *dat* te selecteren, gebruiken we de tag VG(onder;dat). VG staat daarbij voor *voegwoord*, *onder* voor *ondergeschikt*.

Het DPC maakt echter geen onderscheid tussen verschillende types voegwoorden en tussen het voegwoord *dat* en een betrekkelijk voornaamwoord *dat*. Daarom hebben we alle geselecteerde zinnen manueel gecontroleerd. Ter illustratie, in zin 54 is *dat* een betrekkelijk voornaamwoord en geen onderschikkend voegwoord.

- (54) Een reclamebureau in Brussel **dat** onze internetdiensten zou willen integreren in zijn aanbod voor klanten

Verder elimineren we ook zinnen waarbij het voegwoord *dat* deel uitmaakt van een voegwoordelijke uitdrukking⁵ (zie Appendix B voor een overzicht), zoals voorbeeld 55, of waarbij het voegwoord *dat* deel is van een bijzin met graadaanduidend gevolg, zoals in zin 56.

- (55) Omdat de pers dat verlies toeschreef **aan het feit dat** ik mijn verantwoordelijkheden aan mijn schoonzoon overdroeg, besloot ik dat het zo moest zijn.
- (56) Bovendien was ons bedrijf **zo klein dat** we ons konden aanpassen aan de voorwaarden van het contract.

3.1.2.3 De VzC moet een zinsdeel zijn, geen zinsdeelstuk

Een volgend selectie criterium is dat we enkel focussen op VzC's in de vorm van zinsdelen. Het idee is dat de zinsdeelstatus eveneens een effect kan hebben op de positie van VzC's (een effect waar we voorlopig niet in geïnteresseerd zijn). Daarom hebben we in onze dataset een onderscheid gemaakt tussen zinsdelen en zinsdeelstukken. Omdat zinsdelen

⁵ Het gaat hierbij enkel om de selectie van de *dat*-bijzin. VzC's die gevolgd worden door een voegwoordelijke uitdrukking (zie ook Appendix B) nemen we wel op in onze dataset. Een voorbeeld van dat laatste is onderstaande zin. Het werkwoord *focussen* roept het voorzetsel *op* op (o.a. Voorzetsels Prisma) en het voorzetsel wordt gevolgd door een voegwoordelijke uitdrukking *het feit dat*. De VzC staat in een *dat*-zin waarbij *dat* een grammaticaal onderschikkend voegwoord is. Zowel de zin als de VzC beantwoordt aan de selectiecriteria en is bijgevolg relevant.

Ik zou zeggen dat ze niet mogen focussen **op het feit dat ze het als vrouw moeilijk zullen krijgen**.

vrijer kunnen bewegen – een zinsdeel kan in het voorveld staan, wat een zinsdeelstuk per definitie niet kan – gaat onze voorkeur uit naar zinsdelen in de vorm van een VzC. Zinnen met enkel een zinsdeelstuk in de vorm van een VzC werden daarom niet meegenomen in onze uiteindelijke dataset.

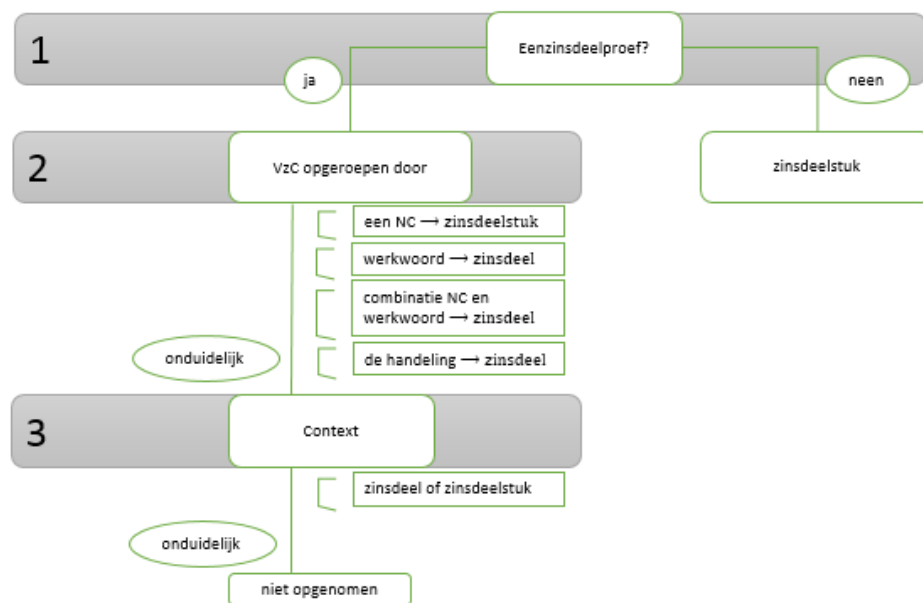
Gesplitste VzC's met een deel in het middenveld en een nabepaling in het achterveld werden in deze studie niet opgenomen. Hoewel de mogelijkheid om te splitsen een ideale manier is om de spanning van een lang middenveld te verlichten, werd het slechts in 24 zinnen geattesteerd. Een voorbeeld is zin 57. De VzC *voor een gemeenschappelijke cultuur die hier en daar een geest kan openen* staat voor een stuk in het middenveld en een stuk in het achterveld. Alle relevante zinnen bevatten m.a.w. een zinsdeel in de vorm van een VzC die in het middenveld óf in achterveld staat.

- (57) Hij gelooft dat hiphop **voor een gemeenschappelijke cultuur** kan zorgen **die hier en daar een geest kan openen**.

Het onderscheid zinsdeel versus zinsdeelstuk brengt echter ook enkele moeilijkheden met zich mee. Zo is het niet altijd duidelijk of een VzC een zinsdeel, dan wel een zinsdeelstuk is. Ter illustratie, in zin 58 kan men twijfelen over de zinsdeelstatus van *in Gibraltar*. Er zijn twee mogelijkheden: ofwel staat de plaats Gibraltar centraal en is de VzC *in Gibraltar* een zinsdeel (een bijwoordelijke bepaling van plaats); ofwel zijn er verschillende types neanderthalers en gaat het in deze zin over de neanderthaler in Gibraltar als een contrast met neanderthalers uit een andere streek. In dat laatste geval is de VzC *in Gibraltar* een zinsdeelstuk (een nabepaling). Zin 58 is met andere woorden syntactisch homoniem.

- (58) Dat wijst erop dat de neanderthalers **in Gibraltar** niet de laatsten van hun soort waren.

Om te bepalen of een VzC een zinsdeel dan wel een zinsdeelstuk is, passen we enkele proeven toe. Deze proeven stellen we stapsgewijs voor in onderstaande beslissingsboom (figuur 7).



Figuur 7: Beslissingsboom zinsdeel versus zinsdeelstuk

De eerste proef is de eenzinsdeelproef. Volgens de eenzinsdeelproef kan enkel een zinsdeel in het voorveld staan en geen zinsdeelstuk. Daarom herformuleren we de bijzin tot een hoofdzin en plaatsen we de VzC in het voorveld. Is het resultaat van deze eenzinsdeelproef ongrammaticaal of nonsensicaal, dan kunnen we de VzC als een zinsdeelstuk beschouwen (ter illustratie zin 59 b).

- (59) a. Dit impliceert dat het proces **van globale opwarming** veel sneller gebeurt dan een paar jaar geleden was gedacht.
b. ***Van globale opwarming** gebeurt het proces veel sneller dan een paar jaar geleden was gedacht.

Lukt het echter wel om een VzC in het voorveld te plaatsen, dan hebben we een eerste indicatie dat de VzC (waarschijnlijk) een zinsdeel is. Enige voorzichtigheid is echter nodig, want de eenzinsdeelproef is niet feilloos. Zowel in zin 58 als in zin 60 kan de eenzinsdeelproef geen sluitend antwoord geven op de vraag of de VzC's *in Gibraltar* en *voor zijn opinies en zijn werk* zinsdelen of zinsdeelstukken zijn. Ze kunnen allebei in het voorveld staan en kunnen tegelijkertijd allebei een naamwoordelijke constituent specificeren, respectievelijk *de neanderthalers* (in zin 58) en *uw waardering* (in zin 60):

- (58) a. Dat wijst erop dat de neanderthalers **in Gibraltar** niet de laatsten van hun soort waren.
b. **In Gibraltar** waren de neanderthalers niet de laatsten van hun soort.
c. **De neanderthalers in Gibraltar** waren niet de laatsten van hun soort.

- (60)
- a. En hij heeft er behoefte aan dat u uw waardering uitspreekt **voor zijn opinies en zijn werk**.
 - b. **Voor zijn opinies en zijn werk** spreekt u uw waardering uit.
 - c. **Uw waardering voor zijn opinies en zijn werk** spreekt u uit.

We moeten m. a. w. nog andere proeven toevoegen naast de eenzinsdeelproef. Een tweede proef is om na te gaan of de VzC opgeroepen wordt door het werkwoord, het gezegde, een samengesteld predicaat of door een constituent. In de eerste drie gevallen spreken we over een zinsdeel. In het laatste geval is de VzC een zinsdeelstuk, namelijk een bijvoeglijke nabepaling of een bijwoordelijke nabepaling. Indien er twee opties zijn, maken we gebruik van drie woordenboeken op basis waarvan we proberen vast te stellen of het werkwoord, een constituent, dan wel een combinatie van beide (i.e. een samengesteld predicaat) de VzC oproept. De drie woordenboeken die we gebruiken, zijn *Van Dale Groot woordenboek van de Nederlandse taal* (het onlinewoordenboek), het voorzetselwoordenboek *Voorzetsels* van Prisma en het *Combinatiewoordenboek* van de Kleijn. Van Dale is het verklarende woordenboek van de Nederlandse taal; *Voorzetsels* Prisma geeft een overzicht van woorden (zelfstandige naamwoorden, adjectieven en werkwoorden) die vaste voorzetsels oproepen of die vaak in combinatie met een bepaald voorzetsel voorkomen; het *Combinatiewoordenboek* van de Kleijn is een lijst van zelfstandige naamwoorden met de bijbehorende vaste werkwoorden. In *Voorzetsels* Prisma staat bijvoorbeeld dat *waardering* gecombineerd kan worden met de voorzetsels *van* of *voor* maar dat deze voorzetsels weggelaten kunnen worden. Volgens Van Dale is *zijn waardering uitspreken over iets of iemand* een vaste combinatie of uitdrukking. Het *Combinatiewoordenboek* van de Kleijn vermeldt tot slot *waardering* in combinatie met *uitspreken*. Als één van de naslagwerken aangeeft dat het voorzetsel, en dus de VzC, door een zelfstandig naamwoord (of adjectief) in combinatie met een werkwoord wordt opgeroepen, beschouwen we de VzC als een zinsdeel. In het geval van zin 60 vinden we *zijn waardering uitspreken over iets* terug in Van Dale en is *voor zijn opinies en zijn werk* een zinsdeel. De attestatie is met andere woorden relevant en wordt opgenomen in dit onderzoek.

Vermeldt geen enkel naslagwerk een combinatie van werkwoord-zelfstandig naamwoord-voorzetsel, maar enkel dat het zelfstandige naamwoord een voorzetsel oproept, dan is de VzC een zinsdeelstuk en wordt het bijgevolg niet opgenomen in dit onderzoek. Een voorbeeld van dat laatste is zin 61. Volgens *Voorzetsels* Prisma kan het woord *actie* met het voorzetsel *tegen* gecombineerd worden en zowel Van Dale als het *Combinatiewoordenboek* van de Kleijn vermeldt *actie ondernemen*. Geen enkel woordenboek vermeldt echter *actie ondernemen tegen*. Daarom beschouwen we in zin 61 de VzC *tegen een drievoudig probleem: het hoge verkrachtingscijfer, het kleine aantal veroordelingen en de seksistische schuldcultuur* als een zinsdeelstuk dat enkel door het zelfstandige naamwoord *actie* wordt opgeroepen en niet door de combinatie *actie ondernemen tegen*.

- (61) De regering moet ingaan tegen deze seksistische schuldcultuur en dit onderzoek beschouwen als een alarmsignaal dat er dringend actie moet worden ondernomen **tegen een drievoudig probleem: het hoge verkrachtingscijfer, het kleine aantal veroordelingen en de seksistische schuldcultuur.**"

In het geval van zin 58 kunnen we geen beroep doen op deze tweede proef en op de drie naslagwerken (Van Dale, Voorzetsels Prisma en het Combinatiewoordenboek van de Kleijn). Daarom hanteren we nog een derde proef, namelijk de context. Op basis van de woorden in de zin zelf en de context voor en na de bestudeerde zin proberen we af te leiden of de onderzochte VzC *in Gibraltar* een plaatsbepaling is, dan wel een nabepaling bij *neanderthaler*.

Context voor: De neanderthalers vonden te weinig water en de dieren waarop ze jaagden, kwamen om. Het was de ergste ramp die de streek de laatste 250.000 jaar had getroffen. Die bevindingen komen er net op het moment dat er een nieuw wetenschappelijk debat op gang is gekomen omtrent de gebieden waar de laatste neanderthalers hebben gewoond. In een andere studie, die is gepubliceerd in het Spaanse Geobios, lieten de onderzoekers van de universiteit van Murcia weten dat ze in een grot in Carihuela, in het noordoosten van Spanje, sedimentlagen hebben ontdekt. Daarin vonden ze werktuigen die dateren van 45.000 tot 21.000 jaar geleden.

Context na: Professor José Carrion stond aan het hoofd van het team. XXXXX Hij wil geen voorbarige conclusies trekken: "De menselijke beenderen zijn tijdens verschillende opgravingen in een periode van vijftig jaar gevonden. Het onderlinge verband en de data die ik opgeef, moeten met het nodige voorbehoud worden behandeld." Professor Finlayson wijst erop dat de data van Carihuela na ijking wel eens zouden kunnen overeenstemmen met die van Gibraltar. Onze eigen soort, de homo sapiens, is ontstaan in Afrika en heeft zo'n 40.000 jaar geleden uiteindelijk de neanderthalers vervangen toen ze Europa is binnengekomen.

Op basis van de voorcontext zouden we geneigd zijn om de VzC *in Gibraltar* als een plaatsbepaling te beschouwen. Het gaat namelijk over de gebieden waar de laatste neanderthalers hebben gewoond. Op basis van de nacontext zouden we eerder kunnen afleiden dat in zin 58 het type neanderthaler bedoeld wordt, namelijk de *neanderthaler in Gibraltar* en niet de *neanderthaler in Carihuela*, zoals het in de context na ook over *de data van Carihuela* en *de data van Gibraltar* gaat. Blijft de zinsdeelstatus van een VzC onduidelijk na het toepassen van deze drie proeven, zoals in zin 58, dan wordt de zin niet opgenomen. We beseffen dat we met deze regel de dataset in zekere mate manipuleren en dat we sommige zinnen misschien wel onterecht elimineren. Tegelijkertijd zijn we op deze manier wel zeker dat alle zinnen die in deze studie als relevant worden beschouwd perfect met elkaar te vergelijken zijn. Zo kan de zinsdeelstatus geen rol spelen in onze dataset en kan het ook de variabele positie van VzC's niet beïnvloeden.

3.1.2.4 De VzC kan zowel in het middenveld als in het achterveld staan

Om de variabele positie van VzC's te onderzoeken, moeten de VzC's in deze studie vanzelfsprekend verplaatsbaar zijn. Dit selectie criterium verlangt dus dat de VzC zowel in het middenveld als in het achterveld kan staan. Volgens de ANS zijn zo goed als alle zinsdelen in de vorm van een VzC verplaatsbaar (ANS 1997: 1372).

Toch zijn er ook VzC's op te sommen die enkel in het middenveld kunnen staan en die we bijgevolg niet meenemen in ons onderzoek. De ANS spreekt over inherente bepalingen. Voorbeelden zijn het naamwoordelijk deel van het gezegde (zin 62) en het niet-werkwoordelijk deel van een uitdrukking (zin 63).

- (62) We hebben alleen willen aantonen dat het Italiaanse voetbal **van een zeer hoogstaand niveau** is.
- (63) Een tweede voorwaarde voor succes is dat het kaf **van het koren** gescheiden wordt.

Een apart verhaal zijn de noodzakelijke richtings- en plaatsbepalingen. Volgens de ANS zijn deze bepalingen inherent en moeten ze daarom in het middenveld staan, net voor de werkwoordelijke eindgroep. Tegelijkertijd stelt de ANS dat bepaalde plaats- en tijdsbepaling als een bijkomstigheid in het achterveld kunnen staan. Het werkwoord of een ander element in het middenveld krijgt dan het accent (zie ook §2.1.1). Het is met andere woorden mogelijk om de VzC *naar Parijs* in zin 64 in het achterveld te plaatsen, zolang de schrijver in dat geval een accent op het werkwoord *komen* legt.

- (64) Vooraanstaande leden van de Palestijnse Autoriteit kondigden aan dat ze **naar Parijs** zouden komen.

Dat een inherente richtings- en of plaatsbepaling in het achterveld kan staan, blijkt ook in onze dataset. In zin 65 bijvoorbeeld roept het werkwoord *samenkomen* een plaats op (*iemand komt ergens samen*) en deze plaats staat in het achterveld.

- (65) De tijd dat enkele verwoede hockeyfans samenkwamen **op een terrein** en daar op vrijwillige basis - en soms vrij amateuristisch - aan het werk gingen, ligt ver achter ons.

Om te bepalen wanneer een werkwoord een accent kan krijgen, waardoor een inherente richtings- of plaatsbepaling in aanmerking komt om in het achterveld te verschijnen, hebben we zelf een extra proef toegevoegd. Deze proef werkt als volgt: als de imperatiefvorm van het werkwoord zonder het inherente complement resulteert in een grammaticaal correcte constructie, dan kan de inherente bepaling in het achterveld staan. In dat geval heeft het werkwoord voldoende betekenis en kan deze betekenis een accent krijgen. Ter illustratie, de imperatiefvormen van zin 64: *Kom!* en van zin 65: *Kom samen!* zijn grammaticaal correct en de werkwoorden kunnen bijgevolg accent krijgen (in zin 64, bijvoorbeeld, kan de komst van de leden van de Palestijnse Autoriteit informatief

belangrijker zijn dan de plaats waar ze naartoe komen). Dezelfde proef geeft echter een ongrammaticaal resultaat bij het werkwoord *belanden* in zin 66:

- (66) Niet meteen de ideale opvolger voor een roman, nee, en als ik had geweten dat ik **op de Booker-longlist** zou *belanden*, had ik het wellicht niet geschreven.

Het werkwoord roept net zoals het werkwoord *komen* een noodzakelijke aanvulling op (iemand/iets belandt ergens), maar kan in tegenstelling tot *komen* niet in de imperatiefvorm zonder complement verschijnen: **Beland!* Volgens onze proef kan het werkwoord *belanden* geen accent krijgen en staat de bepaling *op de Booker-longlist* verplicht in het middenveld. Omdat de VzC niet verplaatsbaar is, wordt zin 66 niet opgenomen.

Vaste uitdrukkingen zoals *aan de lopende band*, *bij uitstek*, *in de eerste plaats*, *in principe*, *in oorsprong*, *naar hartenlust*, *op zijn minst* (of *ten minste*), *op voorhand*, *op zich*, *tot nu toe* (of *tot nog toe*), *van oudsher*, *van tevoren* (of *bij voorbaat*), *voor (het) eerst* en *zonder meer* zijn vergelijkbaar met de inherente bepalingen. Ze kunnen niet in het achterveld staan (zin 67), tenzij ze afgescheiden worden door een intonatiebreuk (zoals in zin 67.c), maar in dat laatste geval behoort de VzC tot de uitloop en niet tot de kernzin (zie ook Broekhuis & Corver 2016)⁶. Net omdat ze enkel in het middenveld voorkomen, nemen we ze niet op in onderhavige studie.

- (67) a. Dit houdt in dat **in principe** één Vlaams deelproject wordt gesteund van een van de best gerangschikte aanvragen, steunend op de internationale peer review en prioriteit.
b. **Dit houdt in dat één Vlaams deelproject wordt gesteund **in principe** [...]*
c. Dit houdt in dat één Vlaams deelproject wordt gesteund, **in principe toch**.

Het valt bovendien op dat deze vaste uitdrukkingen in heel wat gevallen gemakkelijk vervangen kunnen worden door een bijwoord (*aan de lopende band* door *voortdurend*, *in de eerste plaats* door *vooreerst*, enz.) en dat er geen paradigmatische aanpassingen (weglating, uitbreiding, vervanging) mogelijk zijn. In zin 67 zorgt een toevoeging van een definitief lidwoord (*in het principe*) of een aanwijzend voornaamwoord (*in dat principe*) voor een betekenisverschuiving. De oorspronkelijke betekenis, namelijk *zonder rekening te houden met de consequenties of met bijkomende zaken* (uit Van Dale), verdwijnt. In een zin met de VzC *in het/dat principe* lijkt het alsof in een zeker, nog nader te specificeren principe (of grondregel) een project wordt gesteund.

Om te bepalen of een VzC een vaste uitdrukking is met het karakter van een bijwoord, gebruiken we enerzijds de paradigmaproef, waarbij we iets weglaten, toevoegen of vervangen, en anderzijds Van Dale. Van Dale geeft in heel wat gevallen het label *uitdrukking* aan de hogergenoemde combinaties, zoals dat het geval is bij de uitdrukking

⁶ Broekhuis & Corver verwijzen naar de VP-topicalisatieregels om te verklaren waarom deze uitdrukkingen niet in het achterveld kunnen staan (zie voor meer informatie Broekhuis & Corver 2016 en Willems 2016).

aan de lopende band, of geeft een synoniem in de vorm van een bijwoord. Bij de combinatie *op zijn minst* vermeldt Van Dale het bijwoord *minstens* en bij *in oorsprong* het bijwoord *oorspronkelijk*.

Een bepaling van gesteldheid en een voorzetselvoorwerp zijn volgens de ANS ook voorbeelden van inherente bepalingen, behalve als ze in de vorm van een VzC verschijnen (ANS 1997: 1352). In die laatste vorm kunnen ze zowel in het middenveld als in het achterveld verschijnen. Ter illustratie, in zin 68 en 69 kan de VzC, respectievelijk een bepaling van gesteldheid en een voorzetselvoorwerp, zowel in het middenveld als in het achterveld staan. Deze zinnen worden uiteraard opgenomen in deze studie.

- (68) a. Want als de Spaanse voetbalbond tegen de betrokken clubs sancties zou nemen, kan het betekenen dat Osasuna, dat vorig seizoen vierde werd, alsnog **tot kampioen** wordt gekroond.
b. Want als de Spaanse voetbalbond tegen de betrokken clubs sancties zou nemen, kan het betekenen dat Osasuna, dat vorig seizoen vierde werd, alsnog wordt gekroond **tot kampioen**.
- (69) a. Onderzoekers beweren dat het bespieden via de zwarte doos **tot veiliger rijgedrag** leidt.
b. Onderzoekers beweren dat het bespieden via de zwarte doos leidt **tot veiliger rijgedrag**.

Tot slot verlangt dit selectie criterium ook dat de VzC tot de kernzin behoort (de VzC staat namelijk in het middenveld of in het achterveld). VzC's in de uitloop behoren niet tot de kernzin en nemen we bijgevolg niet mee in ons onderzoek (zie zin 70). Deze VzC's worden meestal voorafgegaan door een intonatiebreuk (in geschreven teksten weergegeven door een komma).

- (70) Stel u even voor dat al die opportuniteiten zomaar voor het grijpen liggen, **op nauwelijks enkele meters van uw kantoor**.

Bijstellingen (of tussenvoegingen) in het middenveld zijn vergelijkbaar met de uitloop. Ze zijn bovendien gemakkelijk te herkennen, omdat ze door komma's of gedachtestreepjes afgescheiden worden (zoals in zin 71). Net zoals zinnen met enkel een VzC in de uitloop nemen we ook zinnen met enkel een VzC in een bijstelling niet mee in onze studie.

- (71) Zorg ervoor dat de overheid - **naast het direct stimuleren van het fundamenteel onderzoek** - ook indirect innovatie stimuleert (bijvoorbeeld via de aanpak van intellectuele eigendom, startersbeleid, enz...).

3.1.2.5 De VzC wordt niet voorafgegaan door een focuspartikel en is geen voornaamwoordelijk bijwoord, geen beknopte bijzin, geen *aan het + infinitief*-constructie

We houden enkel rekening met VzC's die met een voorzetsel⁷ of voorzetseluitdrukking beginnen. Dat betekent dat we voorzetsel-achterzetselconstituenten (of VzAzC's) elimineren, zoals *over alle beleidsdomeinen heen* in zin 72, maar evengoed VzC's die door een voorbepaling (zin 73) of door een focuspartikel (zin 74 en 75) worden voorafgegaan.⁸ Door deze selectiecriteria in te voeren, kunnen we het aantal beïnvloedende factoren en het aantal data beperken.

- (72) Dergelijke fenomenen bewijzen ook aan dat beleidsmakers best **over alle beleidsdomeinen heen** denken.
- (73) Als we ons sociale leven van hier vergelijken met dat van België, dan moeten we vaststellen dat we **hier in Polen** veel vaker uitgaan.
- (74) Meer werken voor bedrijven zou betekenen dat we **ook tijdens de week** meer opdrachten krijgen.
- (75) Ik heb daaruit geleerd dat ik **alleen op mezelf** kan rekenen.

Om uit te maken of een focuspartikel bij de VzC zelf staat, dan wel bij het werkwoord, gaan we eerst na of het focuspartikel verplaatsbaar is. Net zoals bij de eenzinsdeelproef herformuleren we de bijzin tot een hoofdzin en plaatsen we het focuspartikel in het voorveld. In zin 75, bijvoorbeeld, kan enkel de VzC met focuspartikel vooraan staan:

- (75) a. **Alleen op mezelf** kan ik rekenen.
b. ***Alleen** kan ik rekenen op mezelf.

⁷ We merken op dat het voegwoord *als* in het DPC als een voorzetsel wordt gelabeld. Zinnen die net op basis van dit foutieve label werden geselecteerd tellen we niet mee. Ter illustratie:

De lezer zal het ons dan ook niet kwalijk nemen dat wij het prospectief beleid hebben weerhouden **als centraal thema van deze review**.

⁸ Een VzC voorafgegaan door een focuspartikel kan niet zomaar naar het achterveld verplaatst worden, tenzij afgescheiden door een komma. In dat laatste geval maakt de VzC geen deel uit van de kernzin maar van de uitloop.

kwade tongen in Hollywood beweerden dat Goddard in diezelfde periode **ook met Frida** een verhouding had
?dat Goddard in diezelfde periode een verhouding had **ook met Frida**
dat Goddard in diezelfde periode een verhouding had, **ook met Frida**

Bovendien kan zelfs de betekenis van de zin veranderen. In de oorspronkelijke zin had Goddard een relatie met dame X en daarnaast ook een verhouding met Frida. De zin met uitloop doet echter vermoeden dat Goddard een relatie had met dame X, een verhouding met dame Y en ook een verhouding met Frida.

Zin 75 nemen we niet op in ons onderzoek omdat de VzC een focuspartikel als voorbepaling heeft.

Toch geeft deze verplaatsingsproef in sommige gevallen geen sluitend antwoord. Een voorbeeld van dat laatste is zin 76. In deze zin kan zowel het focuspartikel met de VzC (zie 76.a) als het focuspartikel alleen (zie 76.b) in het voorveld staan.

- (76) Velen van hen vinden trouwens dat Iran kernenergie **ook voor wapens** mag aanwenden.
a. **Ook voor wapens** mag Iran kernenergie aanwenden.
b. **Ook** mag Iran kernenergie **voor wapens** aanwenden.

Bij dergelijke twijfelgevallen bestuderen we de context voor de onderzochte zin:

Context voor: We kunnen Iran niet vreedzaam en democratisch maken; we kunnen alleen de voorwaarden helpen creëren waarbinnen de Iraniërs die dingen zelf kunnen realiseren. Twee klokken tikken in Iran: de nucleaire en die van de democratie. Het doel van het Westen moet dan ook zijn de nucleaire klok trager te doen tikken en die van de democratie sneller te doen gaan. Maar het probleem is dat sommige dingen die we zouden kunnen doen om het nucleaire proces te vertragen hetzelfde vertragende effect hebben op de democratisering. Miljoenen Iraniërs die zeer kritisch staan tegenover het theocratische regime en hun schreeuwerige president Mahmoud Ahmadinejad, gaan er wel van uit dat Iran recht heeft op kernenergie voor civiele doeleinden.

De voorcontext eindigt met *kernenergie voor civiele doeleinden*. In de bestudeerde zin wordt het focuspartikel gebruikt om een andere reden meer te accentueren. Op basis van deze voorcontext kunnen we besluiten dat het focuspartikel betrekking heeft op de VzC. Deze zin wordt net om die reden niet opgenomen.

We focussen enkel op de ‘klassieke’ VzC’s (met name VzC’s met een NC, AdjC, AdvC of VzC als relatum) en houden geen rekening met voornaamwoordelijke bijwoorden (zin 77) en met beknopte bijzinnen (zin 78). Hoewel voornaamwoordelijke bijwoorden bijna altijd een VzC als pendant hebben (ANS 1997: 490), zijn het bijwoorden zoals de naam ook doet vermoeden.

- (77) Praat over onderwerpen waarvan u vermoedt dat uw gesprekspartner **er** geïnteresseerd **in** is.
(78) Het nadeel daarvan is dat ze moeite hebben **om hun individuele behoeftes aan bod te brengen**.

Volgens de ANS is het bindwoord *om* in zin 78 geen voorzetsel maar een voegwoord (ANS 1997: 1110). In tegenstelling tot de ANS is voor Vandeweghe (2000) het (logische) bindwoord *om* aan het begin van een beknopte infinitiefzin wel een voorzetsel (Vandeweghe 2000: 334) en ook binnen het DPC zijn deze bindwoorden als voorzetsels gelabeld. Omdat de voorzetselstatus van de bindwoorden van een beknopte bijzin onzeker is, nemen we ze niet mee in dit onderzoek. Bovendien is het volgens het

complexiteitsprincipe niet altijd mogelijk om deze beknopte bijzinnen in het middenveld te plaatsen. Het complexiteitsprincipe zegt namelijk dat “elementen die complexer zijn (dat kan betekenen 'zwaarder', 'sterker' of 'omvangrijker/uitgebreider') in de regel verder naar achteren geplaatst worden dan minder complexe elementen” (ANS 1997: 1245). Het vorige selectiecriteria verlaagt echter dat de VzC verplaatsbaar is. Daar bedoelen we mee dat een VzC die in het middenveld staat, evengoed in het achterveld had kunnen staan en vice versa. In voorbeeld 78 kan de beknopte bijzin enkel in het achterveld staan en is deze niet verplaatsbaar naar het middenveld:

- (78) *Het nadeel daarvan is dat ze **om hun individuele behoeftes aan bod te brengen** moeite hebben.

Ongeacht of het bindwoord nu een voorzetsel of een voegwoord is, kunnen we deze zin op basis van het vorige selectiecriteria niet meenemen in onze studie.

Ook voorzetsels in combinatie met een infinitief die deel uitmaken van de werkwoordelijke eindgroep (de typische constructie *aan het* + infinitief; zie zin 79) worden niet opgenomen. Ze staan namelijk noch in het middenveld, noch in het achterveld.

- (79) Tal van enquêtes wijzen erop dat de voorraadbouw **aan het slinken** is en dat de daling van de bedrijfsinvesteringen nagenoeg achter de rug is.

3.1.3 Het effect van de journalistieke subgenres

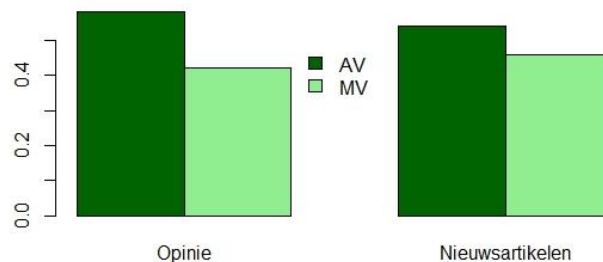
De tekstsoort *journalistieke teksten* bestaat uit verschillende subgenres: opinie (een achtergrondartikel met ruimte voor de mening van de schrijver, een column of een redactioneel geschreven bericht), nieuwsartikelen en informatieve documenten (volgens Delaere (2015: 46) zijn dit “zogenaamde informatieve teksten van algemene aard”, maar noch Delaere, noch de DPC User Manual⁹ geven meer informatie over deze documenten). De eerste twee subgenres zijn het best vertegenwoordigd in onze dataset (1108 zinnen komen uit een opiniestuk, 610 uit een nieuwsartikel en slechts 7 zinnen behoren tot de categorie *informatieve documenten*). Het communicatieve doel van opiniestukken en nieuwsartikelen verschilt in zekere zin en het is niet uitgesloten dat dat een effect kan hebben op de voorkeurspositie van VzC's. Nieuwsartikelen geven voornamelijk een opsomming van de feiten en willen informeren. Ze zijn bijgevolg objectief, of hebben als doel objectief te zijn, en de stem van de journalist is van minder belang. Opiniestukken willen daarentegen de feiten analyseren, evalueren en van commentaar voorzien. De stem van de schrijver is hoorbaar, wat het tekstgenre subjectiever maakt.

⁹ <https://www.kuleuven-kulak.be/dpc/manual/DPC.pdf>

While the main communicative purpose of news reports is basically informative, the goal of news commentaries is analytical, evaluative and persuasive. In functional theories of language, contextual differences –such as the communicative purpose of a given genre– are reflected in the linguistic features used by language producers. (Lavid, Arús & Moratón 2012: 5)

Het subgenre zou dus een effect kunnen hebben op de woordvolgorde van VzC's, een effect dat we bij voorbaat willen controleren. Indien subjectiever zou betekenen dat de mening van de schrijver directer geformuleerd wordt en dat het geschreven taalgebruik meer de gesproken taal benadert, dan zouden we in opiniestukken meer middenveldpositie kunnen verwachten dan in de objectievere nieuwsartikelen. Onderzoek van Jansen (1978) en Braecke (1990) naar de positie van VzC's in gesproken taalgebruik attesteerde namelijk meer middenveldpositie dan achterveldpositie (zie §2.2.4.1).

Om te achterhalen of het subgenre een rol speelt, stellen we de verdeling van de subgenres en de woordvolgordevariatie visueel voor. We gebruiken daarvoor een gecombineerd staafdiagram (zie figuur 8) en stellen vast dat de verhouding middenveldpositie/achterveldpositie in beide subgenres min of meer overeenkomt.



Figuur 8: Staafdiagram van de verdeling middenveld- en achterveldpositie per journalistiek subgenre

We berekenen in een volgende stap de chikwadraatwaarde om na te gaan of het subgenre geassocieerd is met de positie van de VzC (de responsvariabele in dit onderzoek). De nulhypothese veronderstelt dat er geen associatie is, de alternatieve hypothese dat er wel een associatie is:

H_0 : er is geen relatie tussen de twee variabelen; de variabelen zijn onafhankelijk in deze dataset

H_1 : er is een relatie tussen de twee variabelen; de variabelen zijn geassocieerd in deze dataset

De p-waarde is 0,13 en bijgevolg niet significant ($\chi^2 = 2,32$, $df = 1$). We kunnen de nulhypothese m.a.w. niet verwerpen; er blijkt geen relatie tussen de twee variabelen te

bestaan.¹⁰ In deze studie betekent dat dat we geen onderscheid maken tussen de subgenres en dat we de journalistieke teksten als een homogene dataset beschouwen voor de positie van VzC's.

De geselecteerde zinnen (en teksten) zijn allemaal het resultaat van professionele schrijvers, wat ook kan verklaren waarom de woordvolgorde niet significant afwijkt in de verschillende subgenres. Onderzoek naar het schrijfproces van journalisten in verschillende mediavormen en in verschillende tekstgenres heeft aangetoond dat er vooral een verschil bestaat wat de ervaring van schrijvers betreft, meer dan een verschil tussen subgenres. Ervaren schrijvers zouden zich namelijk met meerdere verschillende schrijfstrategieën bezighouden.

Experienced journalists seem much more aware of what they are doing during the writing process: they express process-oriented strategies far more often than inexperienced journalists do. These strategies refer to the writing goal, plan, flow, and monitoring and extend over the writing process as part of a larger project, the actual writing process, and to phases of the writing process. Journalists use these strategies to plan and guide the writing process, for example, to avoid tangents and complete texts on time. Not only do the experienced journalists have more precise ideas of the effect the text dramaturgy should have, they have better strategies at their disposal to recognize and overcome problems while writing. (Perrin & Ehrensberger-Dow 2006: 339)

Over de ervaring van de schrijvers in onze dataset hebben we geen kennis maar we kunnen wel veronderstellen dat ze bepaalde schrijfregels hanteren. Deze schrijfregels hebben ze in hun opleiding of door ervaring geleerd. Het stijladvies om tangconstructies te vermijden (zie ook §2.1.2), zou een van deze schrijfregels kunnen zijn. Indien alle professionele schrijvers met deze 'vermijd-de-tang-regel' rekening houden, dan verklaart dat ook waarom de woordvolgorde niet afwijkt in opiniestukken en nieuwsberichten.

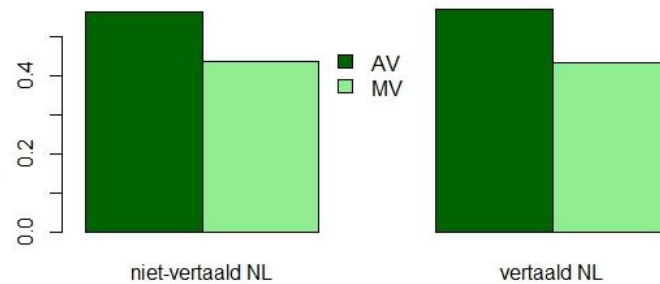
3.1.4 Het effect van vertaald en niet-vertaald Nederlands

Naast verschillende subgenres bestaat het corpus uit zowel vertaald Nederlands als niet-vertaald Nederlands bestaat (een combinatie van vertaald uit het Engels en vertaald uit het Frans). Het idee dat het taalgebruik in vertaalde taal en niet-vertaalde taal anders is, is niet nieuw. Onder anderen Olohan & Baker (2000) stelden vast dat vertaalde taal systematisch verschilt van niet-vertaalde taal. In deze paragraaf gaan we na of de plaatsing van VzC's zich anders gedraagt in deze varianten van het Nederlands. Als dat zo

¹⁰ De verdeling middenveldpositie/achterveldpositie in het ene subgenre is dus niet significant verschillend van de verdeling in het andere subgenre.

zou zijn, dan moeten we in de loop van dit proefschrift de vertaalstatus als extra verklarende variabele opnemen.

In een eerste stap visualiseren we de verdeling middenveldpositie/achterveldpositie voor de vertaalde en niet-vertaalde varianten van het Nederlands (zie figuur 9). Op basis van dit staafdiagram verwachten we niet meteen een significant verschil tussen vertaald en niet-vertaald Nederlands.



Figuur 9: Staafdiagram van de verdeling middenveld- en achterveldpositie in niet-vertaald en vertaald Nederlands

We berekenen de chikwadraattest om statistisch na te gaan of er een associatie bestaat tussen de Nederlandse varianten en de responsvariabele (de positie van VzC's). Net zoals bij de subgenres is de p-waarde niet significant ($p = 0,88$; $\chi^2 = 0,02$ en $df = 1$). We kunnen de vertaalde en niet-vertaalde observaties bijgevolg als een homogene dataset beschouwen.

3.2 Variabelen

Nadat alle zinnen op basis van de genoemde selectiecriteria gefilterd waren, codeerden we de zinnen voor een set onafhankelijke variabelen. Omdat de aard van deze variabelen geen volautomatische codering toeliet, dienden alle attestaties manueel of semi-automatisch gecodeerd te worden. Per variabele werd aan elke zin één waarde toegekend. Op die manier krijgen we een datamatrix met voor elke zin een combinatie van waarden. In de volgende paragrafen stellen we de verschillende onafhankelijke variabelen voor.

De afhankelijke variabele, ook wel de responsvariabele, is de positie van de VzC. Zoals hierboven reeds duidelijk geworden is, houden we rekening met twee mogelijke posities: het middenveld en het achterveld.

3.2.1 De syntactische complexiteit

Net zoals eerder onderzoek naar de positie van VzC's bestuderen we in onderhavige studie ook de *syntactische complexiteit*. Jansen (1978) en Braecke (1990) bestudeerden de *syntactische complexiteit van het middenveld* en de VzC. Zoals we in de inleiding hebben aangehaald, verwachten we dat ook de *syntactische complexiteit van het achterveld* een rol kan spelen. We onderzoeken bijgevolg de *complexiteit* van twee velden, het middenveld en het achterveld, en van de VzC zelf.

De variabele *syntactische complexiteit* operationaliseren we op twee manieren. Enerzijds houden we rekening met de numerieke complexiteit (i.e. de lengte), anderzijds met de hiërarchische complexiteit (i.e. het aantal inbeddingen). De operationalisatie in termen van *lengte* is de meest gebruikte in empirisch onderzoek naar woordvolgordevariatie (Hawkins 1990, Wasow 1997, 2002, Arnold et al. 2000, Szmrecsanyi 2004) en blijkt ook een representatieve manier te zijn om *syntactische complexiteit* te bestuderen (Wasow 1997, Shih 2010). Tegelijkertijd heeft de ANS het over zowel de mathematische uitgebreidheid (de lengte) als de complexiteit (de hiërarchische dieptestructuur) en kunnen we niet a priori veronderstellen dat een langere constituent automatisch een complexere is (eenzelfde redenering vinden we terug bij Jansen 1978; Beaman 1984; Givón 1991; Szmrecsanyi 2004). Daarom houden we rekening met beide operationalisaties, die resulteren in twee variabelen. Voor elke relevante observatie werd dus zowel de *lengte* als het *aantal inbeddingen van het middenveld, de VzC en het achterveld* gecodeerd. De operationalisaties worden verder toegelicht in paragrafen 3.2.1.1 en 3.2.1.2.

3.2.1.1 Toelichting bij de operationalisatie op basis van de *lengte*

De variabele *lengte* hebben we op twee manieren geoperationaliseerd¹¹. Voor alle velden en de VzC tellen we zowel het aantal woorden als het aantal lettergrepen¹². Een verschillende operationalisatie van *lengte* kan nu eenmaal een verschillend resultaat hebben. Om het voorgaande te illustreren, maken we in zin 80 een vergelijking tussen de lengte van het subject (tags: <su> en </su>) en de lengte van de VzC (tags: <vz> en </vz>).

- (80) Het ontstond in 2002 vanuit de vaststelling <begin>dat <su>leidinggevende HR-functies</su> zelden <vz>door een vrouw</vz> <v>worden bezet</v> <end>.

Tellen we het aantal woorden dan stellen we vast dat het subject korter is dan de VzC (2 woorden versus 3 woorden). Tellen we echter het aantal lettergrepen dan blijkt het subject langer te zijn (8 lettergrepen versus 3 lettergrepen). De resultaten (en de implicaties die de resultaten met zich meebrengen) hangen dus af van de operationalisatie van de lengte.

Voor de lengte van het middenveld en voor de lengte van het achterveld houden we rekening met respectievelijk alle woorden tussen de grijpers van de tang en met alle woorden na de werkwoordelijke eindgroep inclusief de uitloop.¹³ De lengte van de relevante en onderzochte VzC tellen we apart (anders wordt deze lengte twee keer geteld: voor de lengte van de VzC en voor de lengte van het middenveld/achterveld). De lengte van het middenveld in zin 81 telt m.a.w. 0 woorden (indien het subject en de persoonsvorm de twee grijpers zijn, zie ook §3.2.1.3), voor het achterveld (met uitloop) 7 woorden en voor de VzC 3 woorden. In het geval de zin gevolgd wordt door een hoofdzin of een nevenschikte bijzin (zoals in zin 81), dan wordt deze laatste niet meegeteld als achterveld (met uitloop).

- (81) De afspraak is <begin>dat <su>de medewerker</su> <vz>voor vier jaar</vz> <v>wordt overgeplaatst</v>, met behoud van de Amerikaanse sociale zekerheid, <end> en dat hij na afloop kan rekenen op een gelijkaardige functie en een aangepaste verloning.

¹¹ Voor het middenveld en het achterveld hebben we ook nog een onderscheid gemaakt tussen het aantal inherente woorden versus het aantal niet-inherente woorden. Bij de niet-inherente woorden stellen we onszelf ook de vraag of deze elementen verplaatst kunnen worden naar het achterveld (voor het aantal niet-inherente woorden in het middenveld) en naar het middenveld (voor het aantal niet-inherente woorden in het achterveld). Daarnaast hebben we ook het aantal zinsdelen geteld in het middenveld. Omdat deze operationalisaties in dit proefschrift niet verder getoetst worden, laten we de bespreking ervan achterwege.

¹² Grondelaers en Speelman 2007 gebruiken in navolging van Arnold et al. (2000) enkel lettergrepen (en geen woorden) om de lengte te operationaliseren: “implementation in terms of number of words would obscure the actual length difference between such pairs as in him and on Madagascar” (Grondelaers & Speelman 2007: 177).

¹³ Rond het subject plaatsen we de tags <su> en </su>, rond de werkwoordelijke eindgroep de tags <v> en </v>. Het einde van de relevante zin sluiten we af met de tag <end>.

Aangezien de *lengte van het middenveld*, *het achterveld* en de *VzC* sterk kan variëren (zie ook tabel 7), zouden we de numerieke continue waarden kunnen groeperen in categorieën.

Syntactische complexiteit				
Operationalisatie		Middenveld	Achterveld	VzC
Lengte	Woorden	0-19	0-36	2-49
	lettergrepen	0-31	0-62	2-87

Tabel 7: Overzicht aantal woorden en lettergrepen voor het middenveld, achterveld en de VzC

We zouden er dus kunnen voor opteren om deze waarden te transformeren tot ordinale variabelen¹⁴. De omzetting van een numerieke variabele naar een ordinale staat echter ter discussie:

A second threat is concerned with whether predictors are used on the most useful information value and scale. As for the former, there is still a lot of work out there in which continuous predictors are factorized. That means, instead of using the continuous predictor as is, researchers break it down into a categorical variable with only a few numbers of levels (maybe by using cut). This can not only lose a lot of information especially if the cutting is not done after a very careful analysis, but it also increases the df for the analysis, potentially making it harder to get significant results. If possible, keeping numeric variables numeric is probably a good idea. (Gries 2013: 254)

In navolging van Gries (2013) werken we enkel met de numerieke variabelen *lengte van het middenveld*, *het achterveld* en de *VzC* in woorden en in lettergrepen.

3.2.1.2 Toelichting bij de operationalisatie op basis van de *inbedding*

De operationalisatie van het *aantal inbeddingen* is sterk gebaseerd op de operationalisatie van complexiteit door Szmrecsanyi (2004). Voor deze variabele tellen we het aantal nabepalingen (en bijstellingen) en houden we rekening met de dieptestructuur in het

¹⁴ De transformatie gebeurt natuurlijk niet arbitrair, maar op basis van een mathematische procedure: de (gepartitioneerde) chikwadraattoets (Agresti 2007: 36). De uiteindelijke onderverdeling van het middenveld, het achterveld en de VzC als ordinale variabelen (op basis van de gepartitioneerde chikwadraattoets) wordt ter illustratie in onderstaande tabel gepresenteerd.

	Categorie1	Categorie2	Categorie3	Categorie4
Lengte middenveld	0 woorden	1-12 woorden	13-19 woorden	
Lengte achterveld	0 woorden	2-36 woorden		
Lengte VzC	2-3 woorden	4-5 woorden	6-11 woorden	12-49 woorden

middenveld, de VzC en het achterveld. Twee complexe types komen daarbij aan bod: een nabepaling (zie voorbeeld 82) en een bijstelling (zie voorbeeld 83).

- (82) <begin>dat <su>administraties en consultants</su> bij het ontwerp **van nieuwe programma's**<v>gebruikmaken</v> <vz>van de informatie **over de opzet van bepaalde types programma's in andere landen**</vz> <end>
- (83) <begin>dat <su>de Mexicaanse president, **Lazaro Cardenas**,<su> asiel <v>gaf</v> <vz>aan Trotski en zijn vrouw, **Natalia**</vz> <end>

De complexiteit van het middenveld, de VzC en het achterveld neemt toe met elke nieuwe ingebedde structuur. Ter illustratie, in zin 82 is er 1 nabepaling in het middenveld (*van nieuwe programma's*), terwijl de VzC uit 3 nabepalingen bestaat (1. *over de opzet*, 2. *van bepaalde types programma's* en 3. *in andere landen*). Het middenveld krijgt daarom waarde 1 voor het *aantal inbeddingen* en de VzC waarde 3. In zin 83 zijn er 2 bijstellingen: *Lazaro Cardenas* en *Natalia*. De eerste eigennaam, *Lazaro Cardenas*, hoort bij het middenveld (het middenveld begint direct na de kern van het subject, hier *president*) en de tweede eigennaam, *Natalia*, maakt deel uit van de VzC. Het middenveld in zin 83 krijgt dus waarde 1 voor *inbedding*, net zoals de VzC.

In zin 84 bevat de VzC een bijzin als nabepaling. Omdat een bijzin de complexiteit misschien wel het meest laat toenemen – “subordinators are the most straightforward indicators of increased embeddedness – and thus of high complexity” (Szmrecsanyi 2004: 1034) – verdubbelen we de waarde. De VzC in zin 84 krijgt dus waarde 2 voor de variabele *aantal inbeddingen*.

- (84) <begin>dat <su>u</su> <v>omringd wordt</v> <vz>door mensen **die uw plannen kunnen uitvoeren en financieren**</vz> <end>

In zin 85 telt het achterveld 2 bijzinnen: een logische voegwoordzin op niveau 1 (als we de ondergeschikte bijzin als niveau 0 beschouwen) en een relatiefzin op niveau 2. De variabele *inbedding van het achterveld* krijgt dus waarde 4 (2 bijzinnen maal waarde 2).

- (85) <begin>dat <su>ze</su> zelfs niet <v>kunnen genieten</v> <vz>van het maal</vz>, **omdat ze spijt hebben van de keuzes die ze niet gemaakt hebben**.

Net zoals bij de variabelen *lengte in woorden* en *in lettergrepen* zullen we ook in het verdere verloop van deze studie enkel met de numerieke variabele *inbedding* werken.

3.2.1.3 Het topologische middenveld versus het cognitief congruerende middenveld

Zoals in het tweede hoofdstuk (§2.1.3) werd aangehaald, kan het middenveld in een Nederlandse bijzin gezien worden als die positie die zich tussen het voegwoord en de werkwoordelijke eindgroep bevindt (d.i. de grammaticale topologische tangconstructie; ANS 1997, Van de Velde 1973) of als die positie die zich tussen het subject van de bijzin en

de werkwoordelijke eindgroep bevindt (d.i. de cognitief congruerende tangconstructie; Uhlenbeck 1962; Jansen 1978; Vrisekoop 2001; Bouman 2009). Om na te gaan of er grote verschillen zijn wanneer de lengte van het middenveld op de ene dan wel op de andere manier geteld wordt, zullen we aan de hand van een correlatietest en een *principal component analysis* nagaan wat de relatie tussen de beide precies is.¹⁵ Het resultaat wordt in hoofdstuk 4, paragraaf 4.1.2.5 besproken.

In het geval van de grammaticale topologische tangconstructie tellen we het aantal woorden tussen het voegwoord en de werkwoordelijke eindgroep (exclusief de lengte van de VzC indien de VzC in het middenveld staat). Voor de cognitief congruerende tangconstructie tellen we het aantal woorden tussen de kern van het subject en de persoonsvorm (exclusief de lengte van de VzC indien de VzC in het middenveld staat). Dat betekent dat eventuele nabepalingen van het subject meegeteld worden, evenals voltooide deelwoorden, infinitieven of partikels die de persoonsvorm voorafgaan. In voorbeeld 86 telt het middenveld volgens de topologische tangconstructie 6 woorden (*er twee twintigers uit Rochdale groot-Manchester*¹⁶) en voor de congruerende tangconstructie 4 woorden (*uit Rochdale groot-Manchester gearresteerd*).

(86) De politie zegt <begin>dat er <su>twee twintigers uit Rochdale, groot-Manchester,</su>
<v>gearresteerd werden</v> <vz>op verdenking van moord</vz> <end>.

3.2.1.4 Hypotheses

Empirisch onderzoek (Jansen 1978, 1979; Braecke 1990; Hoeksema 2014) heeft al vastgesteld dat het middenveld en de VzC de woordvolgorde van VzC's beïnvloeden. We verwachten dus:

1. dat de syntactische complexiteit van het middenveld een positief effect zal hebben op achterveldpositie;
2. dat de syntactische complexiteit van de VzC een positief effect zal hebben op achterveldpositie.

Taalgebruikers zullen bij een complex middenveld en een complexe VzC voor achterveldpositie kiezen, om zo de afstand tussen de grijpers van de tang niet complexer te maken dan nodig en om zo het verwerkingsproces van de lezer te vereenvoudigen (cf. Hawkins 1994; Gibson 2000).

Tegelijkertijd kan ook de syntactische complexiteit van het achterveld een rol spelen. In tegenstelling tot gesproken taal heeft de schrijver de vrijheid om zinsdelen te verplaatsen. Deze vrijheid zou ertoe kunnen leiden dat een schrijver vooral op zoek gaat naar een evenwichtige verdeling in de zin (met zowel zinsdelen in het middenveld als in

¹⁵ Deze twee testen komen ook uitgebreid aan bod in hoofdstuk 4 (meer bepaald in §4.1.2.4 en in §4.1.2.5).

¹⁶ Woorden verbonden door een koppelteken worden als 1 woord geteld.

het achterveld), meer dan naar een lege tangconstructie. Een lezer kan nu eenmaal een zekere afstand tussen de grijpers van de tang overbruggen (zie ook §2.1.2). Hoewel de invloed van het achterveld nog niet bestudeerd werd in eerder empirisch onderzoek naar dit woordvolgordefenomeen, verwachten we:

3. dat de syntactische complexiteit van het achterveld een negatief effect zal hebben op achterveldpositie.

Kort samengevat veronderstellen we een wisselwerking tussen de complexiteit van het middenveld en het achterveld, die eveneens bepaald wordt door de complexiteit van de VzC.

3.2.2 De discoursstatus van de VzC

Naast de *syntactische complexiteit* verwijzen grammatica's (o.a. de ANS) en empirisch onderzoek naar de *discoursstatus* (of *informatiewaarde*) van de VzC als een verklarende factor. De variabele *discoursstatus* komt trouwens ook in andere studies naar andere types van grammaticale variatie in het Nederlands aan bod (o.a., Coussé 2009; Grondelaers & Speelman 2007; Grondelaers et al. 2009; Van Bergen & de Swart 2010 om maar enkele te noemen), en in onderzoek naar alternerende woordvolgordes in het Engels (o.a., Arnold et al. 2000; Gries 2002a; Bresnan et al. 2005; Bresnan & Ford 2010). Er bestaat m.a.w. een rijke onderzoekstraditie naar de informatiestatus van een referent, met verschillende operationalisaties van deze *discoursstatus* tot gevolg.¹⁷

In onderhavige studie werd de variabele *discoursstatus* op verschillende manieren geoperationaliseerd. Ten eerste houden we rekening met de bekendheid van de VzC. Daarvoor operationaliseren we enerzijds de *definietheid* van het relatum van de VzC, en anderzijds de *toegankelijkheid* van het relatum van de VzC. Voor de laatste variabele houden we rekening met de voorcontext en gaan we na of de (belangrijkste) referent(en) van de VzC al eerder vermeld werd(en). Het resultaat is een taxonomie van *maximaal toegankelijke informatie* tot *compleet nieuwe en onverwachte informatie*. Naast de bekendheid bestuderen we ook het belang van de VzC, waarbij we nagaan hoe vaak de referent in de context wordt vermeld. Het achterliggende idee is dat belangrijke informatie meer ter sprake komt dan onbelangrijke informatie. De verschillende variabelen en hun mogelijke levels worden in meer detail in paragrafen 3.2.2.1, 3.2.2.2 en 3.2.2.3 besproken.

¹⁷ Een ander gevolg is dat er verschillende begrippen bestaan om (min of meer) dezelfde lading te dekken. Begrippen zoals *oud*, *thema*, *topic* en *toegankelijk* kunnen meestal vrij door elkaar gebruikt worden (net zoals de antoniemen: *nieuw*, *rhema*, *comment* en *niet-toegankelijk*).

3.2.2.1 Toelichting bij de variabele *definietheid*

De variabele *discoursstatus* wordt heel vaak aan de hand van de definiete (bepaalde) en indefiniete (onbepaalde) determinator van de referent geoperationaliseerd (o.a., Jansen 1978; Gries 2002a; Bresnan et al. 2005; Bresnan & Ford 2010). Definietheid is namelijk het formele middel bij uitstek om nieuwe, onbekende informatie van bekende te onderscheiden (ANS 1997: §14.3.1).

Om de definietheid van de referent (in dit geval de referent van de VzC) concreet te bepalen, gebruiken we het overzicht van Coussé (2009: 6; gebaseerd op de ANS 1997: §14.3.1). Tabel 8 geeft een overzicht van alle determinatoren die een onderscheid maken tussen een definiete en een indefiniete referent.

Definiet	Indefiniet
ANS (1997: §14.3.1):	ANS (1997: §14.3.1):
<ul style="list-style-type: none"> – Definiet lidwoord (<i>de, het</i>) – Possessief pronomen (<i>mijn, jouw,...</i>) – Demonstratief pronomen (<i>die, dit,...</i>) – Relatief pronomen (<i>welke</i>) – Collectief onbepaald pronomen (<i>alle</i>) – Hoofdtelwoord <i>beide</i> 	<ul style="list-style-type: none"> – Indefiniet lidwoord (<i>een, zero</i>) – Vragend pronomen (<i>wie, wat,...</i>) – Niet-collectief onbepaald pronomen (<i>elke, iedere, sommige</i>) – Hoofdtelwoord (<i>een, twee,...</i>)
Toevoeging Coussé (2009):	Toevoeging Coussé (2009):
<ul style="list-style-type: none"> – Eigennamen – Genitiefconstituenten 	<ul style="list-style-type: none"> – Bijzinnen
Eigen uitbreiding:	
<ul style="list-style-type: none"> – Persoonlijke pronomina – Jaartallen – Abstracte nomina 	

Tabel 8: Overzicht van de determinatoren die bepalen of de VzC *definiet* of *indefiniet* is (gebaseerd op Coussé 2009)

Definiete lidwoorden, possessieve en demonstratieve pronomina zijn de determinatoren bij uitstek voor de categorie definiet (net zoals indefiniete lidwoorden voor de categorie indefiniet). In navolging van Coussé (2009) labelen we eigennamen en genitiefconstituenten als definiet en bijzinnen als indefiniet. Coussé argumenteert namelijk dat bijzinnen bij uitstek nieuwe informatie in het discours introduceren. Een eigen uitbreiding betreft de persoonlijke pronomina, jaartallen en abstracte nomina (bijvoorbeeld maanden of dagen, zie ANS 1997: 141). Deze definiete pronomina, hoofdtelwoorden en abstracte nomina worden zelfstandig gebruikt en verwijzen naar een specifiek (bekend) persoon of tijdstip. We labelen ze daarom als definiet.

VzC's met een adjectiefconstituent als complement, zoals zin 87, of een adverbiale constituent, zoals zin 88, krijgen het label NA¹⁸. Het is namelijk onmogelijk om te bepalen of de VzC's *naast stijf en formeel* en *van bovenaf* definitief of indefinitief zijn.

- (87) Het werd het ultieme voorbeeld <begin>dat <su>een pak</su> <vz>**naast stijf en formeel**</vz> ook sexy <v>kon zijn</v> <end>.
- (88) Zijn uitgangspunt is <begin>dat <su>mensenrechten</su> niet <vz>**van bovenaf**</vz> met wetten en formele regelgeving <v>kunnen worden opgelegd</v> indien ze niet van binnenuit gedragen worden door de minder geschoolde bevolking <end>

3.2.2.2 Toelichting bij de variabele *toegankelijkheid*

De onderverdeling definitief versus indefinitief is volgens heel wat onderzoekers een te beperkte operationalisatie van *discoursstatus* (o.a., Givón 1983; Ariel 1990; Grondelaers & Speelman 2007; Grondelaers et al. 2009). Grondelaers en Speelman (2007) stellen bijvoorbeeld vast dat “information marked definite can apparently be newer in some cases than information marked indefinite” (Grondelaers en Speelman 2007: 174). Ter illustratie, in zin 89 uit onze dataset is de VzC *naar het hotel* definitief (de determinator is een definitief lidwoord). We zouden daarom verwachten dat deze VzC oude, bekende en toegankelijke informatie presenteert. Lezen we echter de voorcontext (i.e. tien zinnen voor de geattesteerde zin), dan blijkt het concept *hotel* nog niet eerder vermeld te zijn. Het is met andere woorden nieuwe, ontoegankelijke informatie in de vorm van een definitieve VzC.

- (89) Drie teamleden zijn ervan overtuigd <begin>dat <su>ze</su> hier de rivier <v>moeten oversteken</v> <vz>**naar het hotel**</vz> <end>, maar de vierde besluit een andere weg te nemen omdat hij denkt dat dit geen doorwaadbare plaats is.

Context voor: "Het is donker. Het vierkoppige team dat ik volg, probeert op de kaart uit te zoeken waar ze zich bevinden. Opvallend is dat alle teamleden hetzelfde doen. De taken worden nauwelijks verdeeld. Na bijzonder lang overleg kiezen ze voor de moeilijkste weg: in het pikdonker beginnen ze aan een steile afdaling naar een riviertje. De begeleidende coach roept hen terug omwille van de veiligheid. Nu pas beginnen ze het pad te verkennen. Een teamlid neemt het voortouw, maar doet te veel op eigen houtje. Het duidelijkste teken tot zover, een infobord naast het wandelpad, lopen ze voorbij tot de coach hen erop wijst. Net voor de deadline bereiken ze een plek aan de rivier."

Grondelaers en Speelman (2007), in navolging van Grondelaers (2000), gebruiken daarom een andere classificatie in de vorm van een taxonomie. Deze taxonomie gaat van “maximaal toegankelijke informatie” tot “compleet nieuwe en onverwachte informatie”

¹⁸ In totaal kregen 6 zinnen het label NA voor de variabele *definitetheid*.

(Grondelaers en Speelman 2007: 188). Hoewel de taxonomie oorspronkelijk uit 10 levels bestond, werd die later gereduceerd tot 5 levels: *new*, *discourse-new*, *available in the remote linguistic context*, *available in the recent linguistic context* en *referring to the ground* (Grondelaers en Speelman 2007: 188). Het verschil tussen *nieuwe* en *discours-nieuwe* entiteiten is dat *discours-nieuwe* referenten informatie bevatten die nog niet eerder vermeld is in de voorgaande context, maar die tegelijkertijd wel encyclopedisch toegankelijk is. Een voorbeeld van dat laatste is volgens Grondelaers en Speelman *Bryan Adams* in zin 90 (uit Grondelaers & Speelman 2007: 174):

- (90) Het concert dat **Bryan Adams** op zaterdag 31 mei in Flanders Expo in het kader van zijn “18 till I die”-tour geeft, is al uitverkocht. Er komt een extra show op vrijdag 30 mei.

Hoewel de eigennaam *Bryan Adams* voor het eerst vermeld wordt, veronderstelt de schrijver dat de lezer *Bryan Adams* kent. Voor Grondelaers en Speelman is deze informatie daarom *discours-nieuw*. Deze operationalisatie zorgt wel voor moeilijkheden. Het is namelijk niet eenvoudig om als onderzoeker – en bijgevolg als buitenstaander – te weten of de schrijver uitgaat van enige voorkennis of niet. Deze operationalisatie lijkt me dan ook enkel te werken voor eigennamen zoals in voorbeeld 90.

Ook Arnold et al. (2000) gebruiken een onderverdeling op basis van de toegankelijkheid van een referent. Ze beperken hun hiërarchie tot drie levels: *bekend*, *afleidbaar* en *nieuw*. Informatie is *bekend* als de referent in de eerdere context genoemd wordt en *nieuw* als die niet eerder vermeld wordt. Als de referent niet expliciet vermeld wordt, maar wel uit de context af te leiden is, is die *afleidbaar*.

Voor onze eigen taxonomie gebruiken we eveneens de voorcontext¹⁹, maar baseren we ons op de onderverdeling van Karreman, Enschoot & Schuurs (2013). Zij maken een onderscheid tussen *afhankelijke* en *onafhankelijke verwijzingen*. *Afhankelijke verwijzingen* zijn enkel te begrijpen als de lezer weet naar welk(e) eerdere woord(groep) in de tekst verwezen wordt (2013: 54). In zin 91 bevat de VzC een *afhankelijke verwijzing*. De lezer heeft de voorcontext nodig om te weten wie er bedoeld wordt met het pronomen *haar*.

- (91) Gelukkig weet ze dat ik aan onze toekomst werk en <begin>dat <su>ik</su> echt <vz>**voor haar**</vz> <v>wil knokken</v> <end>."

Onafhankelijke verwijzingen zijn betekenisvolle woorden die naar andere woorden in de tekst verwijzen, maar die de lezer ook zonder die andere woorden zou kunnen interpreteren (2013: 54). Karreman, Enschoot & Schuurs maken wel een onderscheid tussen *lexicale verwijzingen* en *collocationeel lexicale verwijzingen*. *Onafhankelijke lexicale*

¹⁹ De voorcontext bestaat uit maximaal tien zinnen voor de geattesteerde zin met een relevante VzC. We spreken over maximaal omdat een geattesteerde zin ook aan het begin van een journalistieke tekst kan verschijnen. In dat geval bestaat de voorcontext enkel uit een titel.

verwijzingen zijn een letterlijke of een gedeeltelijke herhaling, een synoniem of zelfs een tegenstelling. Ook eigennamen die volledig of gedeeltelijk herhaald worden, en woorden die aan elkaar gerelateerd zijn omdat ze deel uitmaken van dezelfde set (bijvoorbeeld de dagen van de week) zijn *onafhankelijke lexicale verwijzingen*. De *collocationele verwijzingen* doen een beroep op de wereldkennis van de lezer. Een voorbeeld van dat laatste is de VzC *met muziek* in zin 92. In de voorcontext wordt het concept *muziek* niet letterlijk vermeld, maar wel impliciet opgeroepen door het woord *bands* en de verschillende muziekgroepen die opgesomd worden.

- (92) Het siert hem <begin>dat <su>hij</su> al zo lang bezig <v>is</v> <vz>**met muziek**</vz> <end> en dat hij toch niet begint te zeuren dat het vroeger allemaal beter was.

Context voor: Niet omdat het een routine is, maar omdat ik mij niets kan voorstellen dat ik liever zou willen. Ik zou alles kunnen doen wat ik maar wil. Ik heb geen kinderen. XXXXX Ik heb geen banden. Ik zou overal naartoe kunnen gaan en om het even wat doen, binnen de fysieke beperkingen van mijn lichaam. Maar elk jaar vraag ik me af, wat zal ik dit jaar doen? En ik kan niets bedenken dat ik liever wil dan dit. Een van de aantrekkelijkste aspecten van dat immobilisme is zijn eindeloze, frisse enthousiasme. Op de vraag welke *bands* hij goed vindt, rammelt hij een hele waslijst af: *Mogwai, The Cooper Temple Clause, The Rapture, Interpol, Cursive, Thursday, Bright Eyes, Elbow* en nog een stuk of twintig meer. Het zijn er veel en hij zou ze niet kennen als hij niet zo enthousiast was als een tiener.

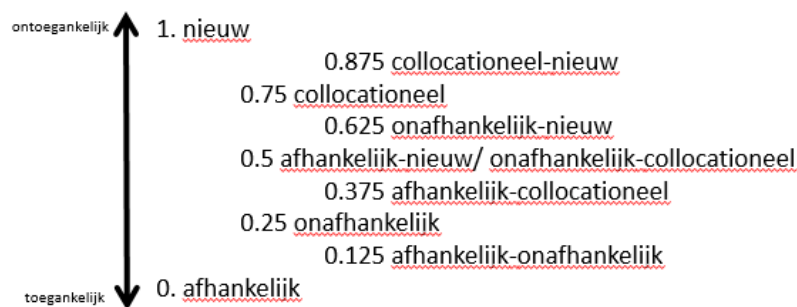
Tot slot is er nog de categorie *nieuwe verwijzingen*, die niet opgeroepen, noch geïmpliceerd worden in de voorcontext. Omdat de relevante VzC's soms meerdere referenten bevatten (vaak nevenschikkend met elkaar verbonden), zijn ze niet altijd in één categorie onder te brengen (met name de vier hoofdcategorieën: *afhankelijk*, (lexicaal) *onafhankelijk*, *collocationeel* of *nieuw*). Daarom voegen we ook combinaties van deze vier categorieën toe. In zin 93 bijvoorbeeld bevat de VzC zowel *afhankelijke informatie* (het possessief pronomen *zijn*) als *onafhankelijk* (*vader* komt ook in de voorcontext ter sprake). Deze VzC wordt als *afhankelijk-(lexicaal)onafhankelijk* gelabeld.

- (93) Hij ontkent <begin>dat <su>hij</su> zijn prestatie in Batman <v>opdraagt</v> <vz>**aan zijn vader**</vz> <end>.

Context voor: Maar ik zou me zorgen maken als ik het opnieuw moest doen." Ook voor Batman was Bale vastberaden. "Ik besloot bij de screentest: 'Ik ga dit doen zoals ik het wil doen en als ze dat niet zien zitten, dan doe ik de film niet'", zegt hij. Hij vreesde dat, zoals in de vorige Warner-films, de slechteriken interessanter zouden zijn dan Batman, terwijl het volgens hem omgekeerd moet zijn. "Is hij gek?", vraagt Bale. "Wat is er aan de hand met die man dat hij midden in de nacht rondrijdt in een vleermuizenpak? Je moet naar een bizarre plaats in je hoofd om dat te aanvaarden. Uiteraard is dat fantasie, maar Chris en ik wilden die vragen beantwoorden met een zo realistisch mogelijke drijfveer." Zoals elke Batman-fan

weet, wordt Bruce Wayne gedreven door de moord op zijn ouders. Dat vindt weerklank bij Bale, wiens eigen vader, David Bale, in december 2003 op 62-jarige leeftijd stierf aan hersenlymfroom.

Figuur 10 stelt onze taxonomie grafisch voor met als uitersten enerzijds maximaal toegankelijke informatie (in onze terminologie: *afhankelijk*; onderaan) en compleet nieuwe en onverwachte informatie (bovenaan):



Figuur 10: Taxonomie van de variabele *toegankelijkheid* (gebaseerd op Karreman, Enschoot & Schuurs 2013)

De twee uiterste levels in figuur 10 krijgen de waarden 0 en 1. De combinatie van deze twee uitersten, namelijk *afhankelijk-nieuw*, krijgt waarde 0,5. We geven de *collocatiele verwijzingen* waarde 0,75 en de *onafhankelijke* waarde 0,25. Alle andere combinaties krijgen een waarde in verhouding tot de hoofdcategorieën (nl. *afhankelijk*, *onafhankelijk*, *collocatieel* en *nieuw*).

Om het aantal levels te beperken, gebruiken we de gepartitioneerde chikwadraattoets. We berekenen met andere woorden de chikwadraatwaarde van het aantal middenveldposities en achterveldposities voor elke mogelijke combinatie (in totaal 45 combinaties of $(n-1)*n/2$). Het resultaat en meer bepaald welke levels we kunnen groeperen, presenteren we in kolommen 4 en 5 van tabel 9. Het valt daarbij op dat *onafhankelijk-collocatiele* en *onafhankelijke verwijzingen* samen clusteren, net zoals *afhankelijk-collocatiele* en *afhankelijk-onafhankelijke verwijzingen*.

Waarde	Oorspronkelijke labels	Waarde	Na chikwadraattoets	Waarde
1	Nieuw	9	Nieuw	8
0.875	Collocationeel-nieuw	8	Collocationeel-nieuw	7
0.75	Collocationeel	7	Collocationeel	6
0.625	Onafhankelijk-nieuw	6	Onafhankelijk-nieuw	5
0.5	Afhankelijk-nieuw	5	Onafhankelijk-collocationeel	4
	Onafhankelijk-collocationeel		Onafhankelijk	
0.375	Afhankelijk-collocationeel	4	Afhankelijk-nieuw	3
0.25	Onafhankelijk	3	Afhankelijk-collocationeel	2
			Afhankelijk-onafhankelijk	
0.125	Afhankelijk-onafhankelijk	2	Afhankelijk	1
0	Afhankelijk	1		

Tabel 9: Variabele *toegankelijkheid*: mathematische waarde (kolom 1); labels (kolom 2), rangorde (kolom 3); labels die significant van elkaar verschillen na gepartitioneerde chikwadraattoets (kolom 4); rangorde na gepartitioneerde chikwadraattoets (kolom 5)

In tegenstelling tot de taxonomie van Grondelaers en Speelman (2007) houden we in onze taxonomie geen rekening met de afstand tussen de verwijzingen. De afstand tussen de laatste verwijzing in de voorcontext en de geattesteerde referent zal zeker de begrijpelijkheid van een tekst beïnvloeden en kan daardoor ook een effect hebben op de woordvolgorde. Een variabele die niet alleen rekening houdt met de toegankelijkheid van de verwijzing maar evengoed met de afstand, zou een terechte uitbreiding zijn in vervolgonderzoek.

3.2.2.3 Toelichting bij de variabele *informationeel belang*

Een tweede manier om *discoursstatus* te operationaliseren is door na te gaan of de informatie die door de VzC gepresenteerd wordt, belangrijk is. Er bestaat namelijk geen correlatie tussen de bekendheid van informatie en het belang ervan. Hoewel nieuwe informatie meer kans maakt om ook belangrijk te zijn, is dit verband niet absoluut, of zoals Grondelaers & Speelman opmerken: “an adjunct need not to be new in order to be the most prominent element” (2007: 185) of ook wel “the correspondence between newness and indefiniteness is too partial to have absolute faith in purely syntactic determinants of informational importance” (2007: 185). De variabele *informationeel belang* is volgens Grondelaers en Speelman (2007) zeer dominant. In die mate zelfs dat ze voorstellen om in sommige contexten de traditionele *oud-voor-nieuw-regel* te vervangen door een *onbelangrijk-voor-belangrijk-regel*:

Since informational prominence appears to overrule newness as an ordering determinant in some contexts, given-before-new had better be rephrased as unimportant-before-important. (Grondelaers & Speelman 2007: 185)

We houden daarom naast *definietheid* en *toegankelijkheid* ook rekening met het *informationeel belang* van de referent. De variabele *informationeel belang* zoals wij die in

onderhavige studie operationaliseren, baseert zich op het principe van “persistence” (Givón 1983). Deze persistentie is “a reflection of the topic’s importance in the discourse” (Givón 1983: 14) en de operationalisatie gaat als volgt:

“persistence is quantified as the number of propositions to the right of element x, in which x “continues an uninterrupted presence as a semantic argument of the clause, an argument of whatever role and marked by whatever grammatical means” (Grondelaers & Speelman 2007: 175).

Een referent die na zijn introductie niet meer vermeld wordt, krijgt zo waarde 0. Een referent die doorlopend aanwezig is in de rest van het discours zal een hogere *persistence*-waarde hebben. Referenten met een hogere *persistence*-waarde zijn informatieel belangrijker dan referenten die maar één keer vermeld worden. Ook Coussé (2009) merkt op dat het belang dat een spreker aan een constituent hecht, uit de rest van het discours blijkt (cataforische referentie): belangrijke constituenten blijven na hun eerste introductie langer in het discours actief dan terloops vermelde constituenten die verder weinig ter zake doen. (Coussé 2009: 4). We breiden Givóns *persistence* uit en berekenen ook hoe vaak de referent in de voorgaande context vermeld wordt (de anaforische referentie). De referent kan namelijk al meermaals ter sprake gekomen zijn in de voorgaande context. Kortom, we tellen voor de variabele *informatieel belang* hoe vaak de referent expliciet (afhankelijk en/of lexicaal onafhankelijk) en impliciet (collocationeel) in de 10 voorgaande en in de 10 volgende zinnen aanwezig is.²⁰ Ter illustratie, in zin 94 komt het woord *patiënten* niet letterlijk ter sprake in de voorcontext. Het wordt echter wel opgeroepen door (of is af te leiden uit) de concepten *psycholoog*, *arts*, *psychiater*, *de behandeling* en *diagnoses stellen/behandelen*. Daarnaast verwijzen de referenten *mensen*, *ze* en *hun* eveneens naar de *patiënten*. We tellen dus 9 verwijzingen in de voorcontext. De nacontext bestaat in dit geval uit 1 zin (einde van het artikel). Het woord *mensen* verwijst opnieuw naar de *patiënten*. De nacontext bevat dus nog één verwijzing en de waarde voor de variabele *informatieel belang*: *nacontext* is 1.

- (94) Professor klinische psychologie Alex Gardner, lid van de British Psychological Society, zegt dat dokters zich er meer van bewust moeten zijn <begin>dat <su>smoorverliefdheid</su> een mogelijke diagnose <v>kan zijn</v> <vz>bij **patiënten**</vz> <end>.

Context voor: Dwangmatig handelen kan ook deel uitmaken van het tot over je oren verliefd zijn. Voorbeelden hiervan zijn het obsessief checken van sms'jes en mails. "De doorsnee klinisch psycholoog zal geen verwijsbrieven krijgen van artsen en psychiaters over smoorverliefdheid", zegt Tallis. "Hoe dan ook zal nauwgezet

²⁰ Deze variabele werd telkens door minstens 2 personen geannoteerd. Bij twijfel of onenigheid werd het gemiddelde genomen van de afwijkende waarden (afgerond naar boven).

onderzoek van het medisch taalgebruik uitwijzen dat dat wel het onderliggende probleem kan zijn." Volgens Tallis hebben al veel mensen om hulp gevraagd omdat ze niet kunnen omgaan met de intensiteit van liefde. Ze zijn er door gedestabiliseerd of lijden eronder dat hun liefde onbeantwoord blijft. In het ergste geval kan dat leiden tot een zelfmoordpoging. Dat is een dramatisering van het oude standpunt dat liefde iemand fataal kan worden, aldus Tallis. Hoewel er tegenwoordig veel onderzoek wordt gedaan naar de behandeling van relationele en psychoseksuele problemen, wordt er nog altijd weinig aandacht besteed aan het probleem van smoorverliefdheid, stelt Tallis. "Misschien is het nu het moment om er serieuzer mee om te gaan en een voorbeeld te nemen aan die oude clinici die diagnoses stelden en behandelden alsof het om gelijk welke klacht ging."

Context na: "Mensen kunnen sterven door een gebroken hart", zegt hij. "Smoorverliefdheid komt waarschijnlijk extreem veel voor."

De variabele *informationeel belang* krijgt in de literatuur ook wel de benaming *zinsfocus* (o.a., Givón 1988; Braecke 1990; Grondelaers & Speelman 2007; Coussé 2009) en bestaat vaak uit een combinatie van *definietheid/toegankelijkheid* met *informationeel belang*. Jansen & Wijnands (2004) bijvoorbeeld combineren *bekendheid* met *belang*. Zo ontstaan er vier mogelijkheden: 1. *bekend en laaginformatief*, 2. *bekend en hooginformatief*, 3. *onbekend en laaginformatief* en 4. *onbekend en hooginformatief*. Coussé (2009) definieert *zinsfocus* als "dat deel van het discours waarin nieuwe informatie geïntroduceerd wordt die belangrijk is voor de spreker" (Coussé 2009: 4). Twee eigenschappen moeten dus door middel van een gedetailleerde discoursanalyse bepaald worden: 1. introduceert de referent nieuwe informatie in de context en 2. kent de spreker prominentie toe aan deze referent. Omdat een dergelijke operationalisatie arbeidsintensief is, maakt Coussé gebruik van de hiërarchie van Givón (1983). De hiërarchie van Givón waarop Coussé (2009) zich baseert, ziet er als volgt uit:

zero > onbeklemtoond / clitisch pronomen > beklemtoond pronomen > definitief
nomen > definitief nomen met een beperkende bepaling

We maken hier twee kanttekeningen bij: ten eerste beperkt de hiërarchie van Givón zich tot definitieve referenten, terwijl ook indefinitieve expressies belangrijk of onbelangrijk kunnen zijn (zie Grondelaers & Speelman 2007; Jansen & Wijnands 2004), ten tweede is deze operationalisatie een versmelting van *definietheid*, *klemtoon* en *complexiteit*²¹. In deze studie wordt de variabele *definietheid* apart geoperationaliseerd (zie 3.2.2.1 voor verdere

²¹ Volgens Givón corresponderen constituenten met het minst fonologische gewicht met informatie die bekend is en waar de spreker weinig aandacht op wil vestigen. Fonologisch langere constituenten corresponderen dan weer met nieuwere of contextueel minder toegankelijke informatie die explicitering nodig heeft (Givón 1988: 249).

toelichting). Bovendien werken we met geschreven teksten, waardoor we niet kunnen horen op welk woord(en) de klemtoon ligt (liggen).

Samengevat houden we de drie operationalisaties van *discoursstatus* gescheiden om zo de rol van elke individuele variabele te vergelijken. Door deze operationalisaties niet te combineren, hebben we evengoed de mogelijkheid om eventuele correlaties en interacties tussen deze variabelen, en dan in het bijzonder tussen de twee operationalisaties van *bekendheid*, vast te stellen.

3.2.2.4 Hypotheses

De regel zegt dat oude informatie voor nieuwe staat, beter bekend als het *given-before-new principle* (Gundel 1988) of het *oud-voor-nieuw-principe*, ook wel het thematische principe (ANS 1997: 1237). Grondelaers en Speelman (2007) stellen echter voor om in sommige contexten de traditionele *oud-voor-nieuw-regel* te vervangen door een *onbelangrijk-voor-belangrijk-regel*. Of we nu over een *oud-voor-nieuw-regel* of een *onbelangrijk-voor-belangrijk-regel* moeten spreken, laten we even in het midden. Wel kunnen we verwachten dat oude/onbelangrijke informatie meer in het middenveld staat en nieuwe/belangrijke informatie meer in het achterveld. Concreet vertaald naar onze operationalisaties *definietheid*, *toegankelijkheid* en *informationeel belang* verwachten we:

4. dat *definiëte* VzC's *middenveldpositie* *verkiezen* en *indefiniëte* *achterveldpositie*;
5. dat *naarmate de* VzC *ontoegankelijker of nieuwer wordt*, *achterveldpositie gestimuleerd wordt*;
6. dat *informationeel* *belangrijkere* VzC's *achterveldpositie* *verkiezen*.

Volgens Arnold et al. (2000) biedt deze *oud-voor-nieuw-volgorde* duidelijke voordelen voor de hoorder en de spreker. Het voordeel voor de hoorder is dat de zin begint met iets dat al gezegd is (en bijgevolg bekend is) en dat zorgt op zijn beurt voor een link tussen wat al gezegd is en wat nog gezegd zal worden. Deze volgorde zorgt dus voor een zekere continuïteit in de tekst, wat de verstaanbaarheid ten goede komt. Daarnaast is deze volgorde ook makkelijker te formuleren. Door de nieuwe referent achteraan te plaatsen krijgt de spreker meer tijd om het complexe proces van de taalhandeling te vervolledigen: de spreker moet eerst beslissen wat hij zal zeggen, vervolgens het idee op een linguïstische manier vormgeven én, zonder veel omhalen, uitspreken. Deze taalhandeling is in een geschreven context minder complex, want de schrijver heeft meer tijd om te formuleren en te herformuleren. Toch blijft het idee dat de *oud-voor-nieuw-volgorde* een link legt tussen de voorcontext en de nieuwe informatie en dat deze volgorde een efficiënte communicatie in de hand werkt, zowel voor een gesproken als voor een geschreven taalhandeling geldig.

Tot slot zouden we volgens het Code-quantity Principle van Givón (1983) kunnen verwachten dat de complexiteit van de VzC ondergeschikt is aan de discoursstatus van de VzC (cf. Braecke 1990). Korte constituenten zouden namelijk corresponderen met

informatie die uit de voorgaande context bekend is en waar de spreker weinig aandacht op wil vestigen. Constituenten met contextueel minder voorspelbare of minder toegankelijke informatie hebben verdere explicitering nodig (Givón 1988: 249) en tellen bijgevolg meer woorden/lettergrepen. De plaatsing van korte, niet-complexe constituenten voor langere, complexere constituenten zou m.a.w. de informatieve geleiding van de zin weerspiegelen.

3.2.3 De grammaticale en semantische status van de VzC's

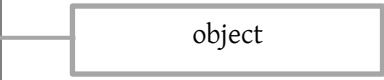
In deze studie operationaliseren we de *grammaticale status* en de *semantische status van de VzC* als twee aparte variabelen. Op die manier kunnen we achterhalen of voornamelijk de grammaticale, structurele functie een rol speelt, dan wel de semantiek, dan wel beide.

Voor de verschillende grammaticale functies die in de vorm van een VzC kunnen voorkomen, hebben we ons gebaseerd op de ANS (1997) en op Vandeweghe (2000). De functies zijn: agensobject (zin 95), bronobject (zin 96), indirect object (zin 97), voorzetselobject (ook wel voorzetselvoorwerp; zin 98), complement (zin 99) en bijwoordelijke bepaling (zin 100).

- (95) Agensobject: Nu heerst de vrees <begin>dat al de archeologische rijkdommen die bewaard liggen in de tot nu toe onaangeroerde wereld, gejat zullen worden <vz>**door genadeloze dieven**</vz> <end>.
- (96) Bronobject: Hebt u het gevoel <begin>dat u steun krijgt <vz>**van de uitgeverijen waarmee u samenwerkt**</vz> <end>?
- (97) Indirect object: Waarmee uw gesprekspartner gewoon bedoelt <begin>dat hij <vz>**aan alle jongeren**</vz> computergames wil verkopen <end>.
- (98) Voorzetselobject: Ik denk trouwens <begin>dat ex-consultants soms iets te sterk focussen <vz>**op cijfers en feiten**</vz> <end>.
- (99) Complement: Volgens de meeste waarnemers is het daarom onvermijdelijk <begin>dat een vrouw <vz>**op de keizerlijke troon**</vz> zal plaatsnemen <end>.
- (100) Bijwoordelijke bepaling: Equicoaching betekent <begin>dat u opdrachten uitvoert <vz>**met een paard**</vz>.

Om de verschillende grammaticale functies van de VzC op een mathematisch onderbouwde manier te groeperen, gebruiken we de gepartitioneerde chikwadraattoets. We berekenen met andere woorden de chikwadraatwaarde van het aantal middenveldposities en achterveldposities voor elke mogelijke combinatie bestaande uit twee grammaticale functies (in totaal 15 combinaties of $(n-1)*n/2$). De functies die samengenomen kunnen worden, i.e. de functies die niet significant van elkaar verschillen wat de verdeling middenveld-/achterveldpositie betreft, presenteren we in kolom 2 van tabel 10.

We merken op dat er geen significante verschillen zijn tussen de plaatsingsvoorkeur van agensobjecten, bronobjecten, indirecte objecten en voorzetselvoorwerpen. Daarom kunnen we deze verschillende grammaticale functies samennemen in een nieuwe categorie *object*. Omdat de literatuur (Jansen 1978; De Schutter 1988; Braecke 1990) het speciale plaatsingsgedrag van voorzetselvoorwerpen benadrukt, nemen we de *voorzetselvoorwerpen* desondanks als een aparte categorie op.

Grammaticale functies	Na gepartitioneerde chikwadraattoets	Nieuwe onderverdeling
Agensobject Bronobject Indirect object Voorzetselvoorwerp		1. object 2. voorzetselvoorwerp
Complement	Complement	3. complement
Bijwoordelijke bepaling	Bepaling	4. bepaling

Tabel 10: Grammaticale functies van de VzC en een nieuwe onderverdeling met behulp van de gepartitioneerde chikwadraattoets

Voor de semantische onderverdeling baseren we ons op de ANS (1997: §20.10). De ANS maakt daarbij het onderscheid tussen bepalingen van: plaats, tijd, frequentie, graad, kwantiteit, maat, causaliteit, gevolg, middel, doel, voorwaarde, toegeving, hoedanigheid, omstandigheid, beperking, verhouding, vergelijking en modaliteit. De betekenissen die wij gevonden hebben in onze dataset, presenteren we in kolom 1 van tabel 11. Drie semantische rollen komen meer dan 100 keer voor: beperking (in categorie 1), plaats (in categorie 2) en tijd (categorie 3 bestaat enkel uit temporele VzC's).

Ook de objecten worden semantisch onderverdeeld: indirecte objecten als doel, agensobjecten als agens en bronobjecten als bron. Een indirect object is een zinsdeel dat “heel algemeen uitgedrukt, aangeeft op wie of wat de werking gericht is” (ANS 1997: 1160). Indirecte objecten bestaan uit meewerkende voorwerpen en belanghebbende voorwerpen (ANS 1997: §20.4). Volgens de ANS zijn belanghebbende voorwerpen “niet altijd duidelijk te onderscheiden van de bepaling van doel”. Vandeweghe is nog explicieter en zegt dat indirecte objecten “het doel van de overdracht noemen” (Vandeweghe 2000: 95). Een bronobject (of een *van*-bepaling) noemt volgens Vandeweghe een bron (vandaar ook benaming bronobject). Een agensobject (of een handelend voorwerp, ANS 1997: §20.8) komt voor in een passieve constructie en correspondeert met het onderwerp (of het agens) in de actieve zin.

Enkel voorzetselobjecten kunnen we niet semantisch labelen. Zij krijgen het label *fixed*, waarmee we verwijzen naar het voorzetsel dat vast opgeroepen wordt door het werkwoord. Deze categorie komt dus helemaal overeen met de categorie *voorzetselvoorwerp* van de variabele *grammaticale status van de VzC*.

Door niet enkel rekening te houden met de klassieke grammaticale functies (objecten, voorzetselvoorwerpen, complementen en bepalingen), maar door de objecten,

complementen en bepalingen ook semantische onder te verdelen, krijgen we dus een nieuwe variabele: de *semantische status van de VzC*²².

We berekenen opnieuw de chikwadraatwaarde van het aantal middenveldposities en achterveldposities voor elke mogelijke combinatie van twee semantische categorieën (in totaal 153 mogelijke combinaties). De nieuwe onderverdeling in tabel 11 is dus niet puur inhoudelijk of arbitrair tot stand gekomen, maar is het resultaat van de gepartitioneerde chikwadraattest.

Semantische categorieën	Nieuwe onderverdeling
Beperking (> 100 attestaties) Bron Agens Causaliteit Doel Frequentie Gevolg Resultaat Toegeving Vergelijking Verhouding Voorwaarde	CATEGORIE 1
Locatief (> 100 attestaties) Hoedanigheid Maat Omstandigheid Middel	CATEGORIE 2
Temporeel (> 100 attestaties)	TEMPOREEL
Fixed (geen semantische rol)	FIXED

Tabel 11: Semantische categorieën van de VzC en een nieuwe onderverdeling met behulp van de gepartitioneerde chikwadraattoets

3.2.3.1 Hypotheses

Voor de grammaticale functie van de VzC verwachten we:

7. dat complementen en objecten middenveldpositie verkiezen en voorzetselvoorwerpen achterveldpositie.

²² We gebruiken in dit proefschrift de benaming ‘semantische status’ met als opmerking dat deze variabele enerzijds uit voorzetselvoorwerpen bestaat en anderzijds uit een semantische herverdeling van de klassieke grammaticale categorieën objecten, complementen en bepalingen.

Volgens de Nederlandse grammatica zijn complementen en objecten inherent verbonden met het werkwoord. We kunnen dus verwachten dat ze de positie voor de werkwoordelijke eindgroep verkiezen (i.e. middenveldpositie). Op basis van het onderzoek van Jansen (1978), De Schutter (1988) en Braecke (1990) kunnen we ook verwachten dat voorzetselvoorwerpen meer in het achterveld staan.

Wat de positie van bijwoordelijke bepalingen betreft, kunnen we geen concrete hypothese vooropstellen. Bepalingen zijn niet inherent verbonden met het werkwoord en zijn daardoor makkelijk verplaatsbaar. Of dat ook betekent dat ze meer in het achterveld staan, is niet op voorhand te bepalen.

Voor de semantische rol van de VzC verwachten we:

8. *dat VzC's van tijd en plaats middenveldpositie verkiezen.*

Eerder onderzoek van Jansen (1978) en Braecke (1990) stelde namelijk vast dat tijds- en plaatsbepalingen meer middenveldpositie verkiezen in vergelijking met andere semantische rollen. In onze studie zou dat betekenen dat categorie 2 (met o.a. de locatieve VzC's) en de temporele VzC's opvallend meer in het middenveld staan in vergelijking met categorie 1 en de *fixed*-groep.

Aangezien de *fixed*-groep volledig overeenkomt met de grammaticale functie *voorzetselvoorwerp* verwachten we voor dat laatste een duidelijke voorkeur voor achterveldpositie (hypothese 7). Strikt genomen is de semantische variabele, *de semantische status van de VzC*, een combinatie van drie semantische categorieën en één grammaticale functie (*voorzetselvoorwerp*).

Jansen (1978) en Braecke (1990) combineerden in hun onderzoek ook de grammaticale en semantische functies in één variabele. Jansen (1978) baseert zich voor zijn variabele *functie* op De Haan (1976) en maakt een onderscheid tussen *VP-bepalingen* (richting, plaatsbepaling en voorzetselvoorwerpen) en *Predicate Phrase bepalingen* (alle andere bijwoordelijke bepalingen). De variabele *functie* bij Jansen bestaat dus uit zowel grammaticale functies van de VzC's – enerzijds bijwoordelijke bepalingen en anderzijds voorzetselvoorwerpen – als uit de semantische rol die de VzC heeft (richting, plaats). Braecke (1990) baseert zich op De Schutter (1976) en houdt met meerdere semantische rollen rekening, namelijk met (i) hoedanigheid, (ii) omstandigheid, (iii) middel, (iv) tijd, (v) plaats, (vi) richting, (vii) restrictie en (viii) reden, oorzaak, doel. Naast deze semantische functies bevat Braeckes variabele ook voorzetselvoorwerpen, een tweede object en een handelend voorwerp.

3.2.4 Het werkwoord

Uit het literatuuroverzicht is al gebleken dat in eerder empirisch onderzoek naar de positie van VzC's de focus op de VzC en het middenveld lag (zie Jansen 1978, 1979, 2003; Braecke 1990; Hoeksema 2014). De grote afwezige in dit verhaal is misschien wel de invloed van het werkwoord. Corpusgebaseerd onderzoek naar Engelse alternanties heeft namelijk al aangetoond dat het werkwoord een rol speelt in de keuze tussen alternerende volgordes (zie o.a., Stallings, MacDonald & O'Seaghdha 1998; Gries 2004; Bresnan et al. 2005; Bresnan & Ford 2010), en een soortgelijke rol verwachten Speelman et al. (2009) ook voor het Nederlandse werkwoord:

While the listener/reader uses the unfolding text/conversation to make predictive inferences about upcoming materials (so that new information can be integrated more easily), the speaker helps the listener/reader by using a conventionalized grammatical format which has an adjunct (On the hook, In the cupboard) and (especially in Dutch) a specific verb in first and second position which strongly constrain the range of possible continuations. This helps the listener to predict new information and to integrate it in the existing context (Grondelaers et al. 2009)

Omdat het werkwoord de zin vormgeeft en de constituenten kadert binnen een bepaalde context, willen we de invloed van het werkwoord op de woordvolgordevariatie van VzC's testen. Volgens Stallings, MacDonald & O'Seaghdha (1998) is het essentieel om de eigenschappen van individuele werkwoorden te onderzoeken en niet zozeer te focussen op overkoepelende werkwoordklassen (Stallings, MacDonald & O'Seaghdha 1998: 413). Door de individuele werkwoorden als een *random factor* in een *Generalized Linear Mixed Model* (Baayen 2008: 278-284) op te nemen, zouden we aan deze voorwaarde beantwoorden. Met behulp van een *Generalized Linear Mixed Model* kunnen we namelijk achterhalen of de individuele werkwoorden überhaupt een rol spelen en of ze een verklarende en voorspellende kracht hebben. Het grote nadeel aan deze methode is dat het model niet zegt op welke manier de werkwoorden de woordvolgordevariatie verklaren en of een bepaald type werkwoord meer achterveldpositie (of middenveldpositie) verkiest.²³ Daarom willen we de zeer diverse en grote groep aan werkwoorden – de 1725 relevante zinnen bevatten in totaal 571 individuele werkwoorden – ook onderverdelen in vooraf bepaalde variabelen (of *fixed factors*).

Voor de eerste onderverdeling van het werkwoord zoeken we de frequentie van elk hoofdwerkwoord op. We proberen ook de werkwoorden semantisch onder te verdelen. Het idee dat de syntactische eigenschappen van een werkwoord vooraf bepaald worden

²³ Stel dat de variabele *individueel werkwoord* veel variatie verklaart en voorspelt, dan willen we uiteraard graag weten op welke manier dat precies gebeurt. Het model definieert echter niet hoe we deze individuele werkwoorden moeten interpreteren; welk type werkwoord heeft een voorkeur voor welke positie?

door de betekenis van het werkwoord, werd al meermaals bestudeerd en erkend (zie o.a. de referentiewerken van Fillmore 1968 en Levin 1995). Tot slot gebruiken we ook de Collostructional Analysis van Gries & Stefanowitsch (2004) om na te gaan welke werkwoorden een statistisch significante voorkeur hebben voor het midden- of het achterveld. Het resultaat is een verdere opsplitsing op basis van het type hoofdwerkwoord en op basis van de zinsrelator. Voorgaande operationalisaties worden in meer detail in de paragrafen 3.2.4.1, 3.2.4.2 en 3.2.4.3 besproken.

3.2.4.1 Toelichting bij de variabele *frequentie van het hoofdwerkwoord*

Voor deze variabele vertrekken we vanuit het idee dat de frequentie van het werkwoord een invloed zou kunnen hebben op de woordvolgorde in de zin. Om deze frequentie te berekenen, kiezen we voor het SUBTLEX-NL-corpus. Brysbaert & New stellen namelijk vast dat ondertitels van films en televisieprogramma's het alledaagse woordgebruik beter benaderen dan geschreven bronnen (Keuleers, Brysbaert & New 2010: 644). Ze raden daarom het SUBTLEX-NL-corpus aan boven elke ander geschreven corpus, zoals het CELEX-corpus:

Compared with the CELEX frequencies for Dutch, the SUBTLEX-NL frequencies are an improvement of almost 10% in explained variance in RTs [reading times]. Therefore, we think that the SUBTLEX-NL word frequencies will be of valuable use for language research in general, and for word recognition research in particular. Although the lexical information contained in the CELEX lexical database remains invaluable, the SUBTLEX-NL word frequencies should be preferred over the CELEX frequencies when selecting stimuli for experiments. (Keuleers, Brysbaert & New 2010: 648)

We gebruiken de filter *all.pos.lemma.freq* wat staat voor “the SUBTLEX-NL lemma frequency of each PoS [i.e. Part of Speech] taken by the stimulus word”. SUBTLEX-NL geeft voor elk hoofdwerkwoord in onze dataset de frequentie (uit een groot, onafhankelijk corpus). Daaruit blijkt, bijvoorbeeld, dat het werkwoord *zijn* de hoogste frequentie heeft, namelijk 1.891.414.²⁴

Tabel 12 geeft tot slot een overzicht van de tien meest voorkomende werkwoorden in onze studie en in het SUBTLEX-NL-corpus. Vijf werkwoorden vinden we in beide corpora terug, wat op zich veel is aangezien het SUBTLEX-NL-corpus ook rekening houdt met de frequentie van hulpwerkwoorden (minder variatie waardoor ze zeer frequent voorkomen) en we in eigen onderzoek enkel focussen op hoofdwerkwoorden. De werkwoorden *maken* en *worden* vinden we ook terug in de top-20 van de meest frequente werkwoorden in het SUBTLEX-NL-corpus.

²⁴ Indien een werkwoord niet voorkwam in het SUBTLEX-NL-corpus dan kreeg het waarde 0.

	Eigen dataset	SUBTLEX-NL
1	Zijn	Zijn
2	Hebben	Hebben
3	Maken	Gaan
4	Gaan	Kunnen
5	komen	Doen
6	staan	Willen
7	doen	Moeten
8	werken	Zullen
9	worden	Weten
10	leiden	Komen

Tabel 12: Overzicht van de 10 meest voorkomende werkwoorden in onze studie en in het SUBTLEX-NL-corpus

3.2.4.2 Toelichting bij de *semantische classificatie van het werkwoord op basis van het Referentiebestand Nederlands*

Het spreekt voor zich dat 571 werkwoorden manueel onderverdelen in semantische klassen niet eenvoudig is. We zouden een beroep kunnen doen op de *verb classes* van Levin (1995), maar ten eerste bevat deze lijst enkel Engelse werkwoorden en geen Nederlandse en ten tweede zijn niet alle werkwoorden uit onze dataset hierin opgenomen (en zeker niet de betekenisarme werkwoorden en koppelwerkwoorden). Daarom gebruiken we een Nederlandstalige lexicale databank met semantische informatie, namelijk het Referentiebestand Nederlands.

Het Referentiebestand is een corpusgebaseerde lexicale databank van het Nederlands met ruim 45.000 trefwoorden en ruim 90.000 voorbeelden. Bij elke betekenis geeft het Referentiebestand gedetailleerde informatie over de orthografie, de morfologie, de syntaxis, de semantiek, de pragmatiek en de combinatoriek. Volgens Levin zijn werkwoordklassen “sets of semantically-related verbs sharing a range of linguistic properties” (Levin 1995: 14). In het Referentiebestand Nederlands komen de werkwoordklassen eveneens tot stand op basis van de aan- of afwezigheid van de kenmerken [*dynamic*], “het dynamische kenmerk kan met de syntactische test *aan het V...en* vastgesteld worden. Anders gezegd impliceert [*dynamic*] verandering, state niet.” (Martin & Maks 2005: 71), en [*control*], “De control-parameter is positief wanneer het subject in de combinatie S(ubject) + V(erb) actief het gebeuren bepaalt, m.a.w. bepaalt of V plaatsvindt of niet.” (Martin & Maks 2005: 71). In totaal ontstaan er zo drie semantische hoofdklassen die we in tabel 13 presenteren:

Semantische klasse	[dynamic]	[control]
ACTION	[+]	[+]
PROCESS	[+]	[-]
STATE	[-]	[+/-]

Tabel 13: Semantische onderverdeling werkwoorden volgens Referentiebestand Nederlands

Het Referentiebestand Nederlands onderscheidt *action* werkwoorden, *process* werkwoorden en *state* werkwoorden.²⁵ Werkwoorden die niet aanwezig waren in het Referentiebestand krijgen waarde NA.²⁶ Het referentiebestand Nederlands voert nog een verdere onderverdeling door op basis van twee bijkomende criteria: 1. De valentie van het predicaat (intransitief, transitief en ditransitief) en 2. het gebruik van attributieve (“the verb expresses a relation of ownership”), spatiale (“the verb expresses a location or movement of (one of) the participants(s)”) of cognitieve (“the verb demands emotional, perceptual or mental activity”) primitieve features (Martin & Maks 2005: 72; Maks, van der Vliet, Görög & Vossen 2013: 11-13). De features en de namen van de categorieën worden gepresenteerd in tabel 14 (uit Maks, van der Vliet, Görög & Vossen 2013: 12-13). De subclassificatie (voorgesteld in kolom 1 en 2 in tabel 14) werd jammer genoeg maar voor enkele werkwoorden uitgevoerd waardoor we zelf een verdere classificatie zouden moet uitvoeren. Door het erg tijdrovende karakter ervan, hebben we voorlopig enkel rekening gehouden met de valentie van het werkwoord. Intransitieve *action*-werkwoorden krijgen zo het label *action 1*, transitieve *action*-werkwoorden het label *action 2* en ditransitieve *action*-werkwoorden *action 3*. Als het werkwoord een complement van plaats (of richting) oproept, geven we het werkwoord het label *action_spatial* (en afhankelijk van de transitiviteit gevolgd door een cijfer 1, 2 of 3). Dezelfde operationalisatie hebben we uitgevoerd voor de *state*- en de *process*-werkwoorden.

Naam	Features	Eigen indeling
Action 1	Control Dynamic	Action1
Action 2	Control Dynamic Transitive	Action2
Action 3	Control Dynamic Ditransitive	Action3
Echprod 2	Control Dynamic Attributive Transitive	
Echprod 3	Control Dynamic Attributive Ditransitive	
Mvmt 1	Control Dynamic Spatial	Action_spatial1
Mvmt 2	Control Dynamic Spatial Transitive	Action_spatial2
Mvmt 3	Control Dynamic Spatial Ditransitive	Action_spatial3
Cognt 1	Control Dynamic Cognitive	
Cognt 2	Control Dynamic Cognitive Transitive	
Cognt 3	Control Dynamic Cognitive Ditransitive	
State 1		State1
State 2	Transitive	State2
State 3	Ditransitive	State3
Possess 2	Attributive Transitive	
Location 1	Spatial	State_spatial1

²⁵ Ook de combinaties *action/process* werkwoorden en *action/state* werkwoorden komen voor in het Referentiebestand Nederlands.

²⁶ Er zijn in totaal 8 zinnen met een waarde NA voor de variabele *semantische classificatie van het werkwoord*. Het gaat meer bepaald over de werkwoorden *achterophinken*, *hervallen*, *meetrillen*, *samenspelen*, *verlokken*, *wegbijten* en *zich positioneren*.

Location 2	Spatial Transitive	State_spatial2
Location 3	Spatial Ditransitive	State_spatial3
Stcognt 1	Cognitive	
Stcognt 2	Cognitive Transitive	
Stcognt 3	Cognitive Ditransitive	
Process 1	Dynamic	Process1
Process 2	Dynamic Transitive	Process2
Prmvmt 1	Dynamic Spatial	Process_spatial1
Prmvmt 2	Dynamic Spatial Transitive	Process_spatial2
Procognt 1	Dynamic Cognitive	
Procognt 2	Dynamic Cognitive Transitive	
Procognt 3	Dynamic Cognitive Ditransitive	

Tabel 14: Kolom 1 en 2 is de subclassificatie volgens Maks, van der Vliet, Görög & Vossen (2013: 12-13); kolom 3 onze eigen semantische (sub)classificatie

3.2.4.3 Toelichting bij de Collostructional Analysis (Gries & Stefanowitsch 2004): de variabele *type hoofdwerkwoord* en *zinsrelator*

Om na te gaan of individuele werkwoorden een andere woordvolgorde prefereren, gebruiken we de *Distinctive-collexeme of Collostructional Analysis* van Gries en Stefanowitsch (2004). Deze analyse bestudeert de individuele werkwoorden en legt de voorkeur van elk werkwoord voor een van de beide volgordevarianties bloot. Het idee is dat we op die manier het bestaan en de graad van semantische verschillen tussen de twee woordvolgordes kunnen tonen:

An analysis of the verbs that are distinctive for each construction (i.e. that distinguish it significantly from the other) may help us elucidate the existence and degree of fine semantic differences between the two that might, in turn, explain the different restrictions. (Gries & Stefanowitsch 2004a: 102)

De analyse zelf gaat als volgt te werk:

In order to calculate the distinctiveness of a given collexeme, we need four frequencies: the lemma frequency of the collexeme in construction A, the lemma frequency of the collexeme in construction B, and the frequencies of construction A and construction B with words other than the collexeme in question. These can then be entered in a 2-by-2 table and submitted to the Fisher exact test (or any other distributional statistic) (Gries & Stefanowitsch 2004a: 102)

Tabel 15 illustreert deze methode. Het werkwoord *betalen* komt 5 keer voor in middenveldpositie en 8 keer in achterveldpositie. Deze aantallen worden tegenover alle andere werkwoorden in middenveldpositie en alle andere werkwoorden in achterveldpositie geplaatst. Vervolgens wordt de Fishers exact test berekend (in het geval van het werkwoord *betalen* is het resultaat niet significant; $p > 0.05$).

	<i>betalen</i>	Andere werkwoorden	Totaal
Middenveldpositie	5	742	747
Achternveldpositie	8	970	978
<i>Totaal</i>	13	1712	1725

Tabel 15: De verdeling van het werkwoord *betalen* tegenover alle andere werkwoorden

Met behulp van de *Collostructional Analysis* krijgen we uiteindelijk een overzicht welke werkwoorden een significante voorkeur hebben. Deze werkwoorden worden in tabel 16 gepresenteerd. We merken hierbij op dat het reflexieve voornaamwoord bij reflexieve werkwoorden, het partikel bij een partikelwerkwoord, het niet-werkwoordelijke deel van een werkwoordelijke uitdrukking en het gezegde van een samengesteld predicaat, mee opgenomen zijn onder de noemer *werkwoord*.²⁷

Werkwoord	
Middenveld	Achternveld
doen	rol_spelen
komen	deel_uitmaken
lijden	recht_hebben
zorgen	verantwoordelijk_zijn
gaan	
beschikken	
spelen	
blijken	
kopen	
praten	

Tabel 16: Werkwoorden die op basis van de collostructional analysis (Gries & Stefanowitsch 2004) een voorkeur hebben voor een van de beide woordvolgordes ($p < 0.05$)

We stellen vast dat de 5 zeer hoogfrequente werkwoorden die onze dataset gemeen had met het SUBTLEX-NL-corpus (*zijn*, *hebben*, *gaan*, *komen* en *doen*) ook in tabel 16 aanwezig zijn. Dat is op zich niet verwonderlijk. De Fishers exact test, die in deze *Collostructional Analysis* gebruikt wordt, berekent het verschil in het aantal absolute attestaties. Werkwoorden die maar twee keer geattesteerd worden (vaak laagfrequente werkwoorden), mogen dan nog resoluut in één positie voorkomen, het gebrek aan attestaties zal ervoor zorgen dat de *distinctive collostructional strength* laag is en dat het resultaat bijgevolg niet significant is.

²⁷ We hadden ook enkel kunnen focussen op het werkwoord, zonder rekening te houden met een reflexief voornaamwoord, een partikel of een uitdrukking. Voor de volledigheid hebben we ook een Collostructional Analysis uitgevoerd voor deze operationalisatie van het werkwoord. De werkwoorden die een significante voorkeur voor middenveldpositie hebben, zijn: *doen*, *zorgen* *beschikken*, *blijken*, *kopen*, *lijden*, *praten*, *vallen* en *gaan*. De werkwoorden die achternveldpositie prefereren, zijn: *hebben* en *zijn*.

Wat op zich wel verwonderlijk is, is de verdeling van deze hoogfrequente werkwoorden over het midden- en achterveld. Een VzC in combinatie met een zelfstandig werkwoord, zoals de werkwoorden *gaan*, *komen* en *doen*, zal vaker in middenveldpositie staan. De werkwoorden *zijn* (meestal een koppelwerkwoord) en *hebben* verkiezen een VzC in achterveldpositie.

We maken een eerste, zeer grofmazig onderscheid op basis van het zinswerkwoord. Het zinswerkwoord is de lexicale kern, “omdat het instaat voor de verbinding met de andere elementen” (Vandeweghe 2000: 46). Er zijn twee mogelijke zinswerkwoorden: een zelfstandig werkwoord en een koppelwerkwoord. Daarnaast maken we een onderscheid op basis van de zinsrelator. Tabel 16 toont namelijk dat hoogfrequente en tegelijkertijd zeer betekenisarme werkwoorden in een uitdrukking of een samengesteld predicaat (bv. *behoefte hebben aan*, *gek zijn op*, enz.) een VzC in achterveldpositie verkiezen. We maken bijgevolg een onderscheid tussen 1. *samengestelde werkwoorden* (dus werkwoorden in een uitdrukking of in een samengesteld predicaat), 2. *enkelvoudige werkwoorden*, 3. *reflexieve werkwoorden* en 4. *werkwoorden met een partikel*. Combinaties van deze verschillende levels zijn eveneens mogelijk, zoals 5. *samengestelde werkwoorden met een partikel*, 6. *samengestelde reflexieve werkwoorden* en 7. *reflexieve werkwoorden met een partikel*. Met behulp van de gepartitioneerde chikwadraattoets stellen we echter vast dat categorieën 3 tot 7 niet significant van elkaar verschillen waardoor we ze in één groep samennemen. Kort samengevat bestaat de factor uit drie categorieën: *samengestelde werkwoorden*, *enkelvoudige werkwoorden* en *werkwoorden met een partikel*, *reflexieve werkwoorden of een combinatie van de voorgaande* (voor een overzicht, zie tabel 17).

De zinsrelator		
Categorie 1	Categorie 2	Categorie 3
Samengestelde werkwoorden	Enkelvoudige werkwoorden	- Reflexieve werkwoorden - Partikelwerkwoorden - Samengestelde werkwoorden met partikel - Samengestelde reflexieve werkwoorden - Reflexieve werkwoorden met partikel

Tabel 17: Onderverdeling variabele *zinsrelator*

3.2.4.4 Hypotheses

In tegenstelling tot de vorige variabelen kunnen we voor de rol van het werkwoord niet naar het empirische onderzoek in paragraaf 2.2.4 verwijzen. De reden is dat het werkwoord als een verklarend mechanisme voor de woordvolgordevariatie van VzC's nieuw is en nog niet eerder onderzocht werd. We kunnen echter wel enkele

verwachtingen formuleren op basis van andere (en voornamelijk Engelse) studies naar het werkwoord en woordfrequentie.

Onderzoek naar tekstverwerking (en de tijd die nodig is om een tekst te verwerken) toont aan dat de woordfrequentie het grootste deel van de variantie verklaart: “In large-scale studies, word frequency (WF) reliably explains the largest percentage of variance of any predictor of word processing times” (Keuleers, Brysbaert & New 2010: 643, zie ook Baayen, Feldman & Schreuder 2006; Balota, Cortese, Sergent-Marshall, Spieler & Yap 2004; Yap & Balota 2009). Het achterliggende idee is dat hoogfrequente woorden makkelijker te onthouden, te herkennen, te formuleren en te verwerken zijn dan laagfrequente woorden. Gibson (1998) en Ellis (2002a) stellen in hun onderzoek eveneens vast dat hoogfrequente woorden makkelijker geactiveerd worden dan laagfrequente woorden (“high frequency lexical items require fewer energy resources to become activated than low frequency lexical items do” (Gibson 1998: 9). Omdat deze hoogfrequente (werk)woorden makkelijker geactiveerd worden, zijn ze ook makkelijker te voorspellen waardoor ze een effect kunnen hebben op de woordvolgorde. Wat dat effect precies is op de positie van VzC’s, is moeilijk op voorhand in te schatten. Wel is het zo dat hoogfrequente woorden heel diverse en daardoor soms heel vage betekenissen dragen (Stahl & Nagy 2006). Deze semantische verbleking zorgt ervoor dat het werkwoord een inherent zinselement nodig heeft dat verplicht in het middenveld staat. Daardoor is het middenveld al behoorlijk ingevuld (door het inherente element en door het subject). In het geval de schrijver het middenveld niet (nog) langer of complexer wil maken, zouden we kunnen verwachten dat het achterveld de voorkeur geniet bij hoogfrequente of semantisch verbleekte werkwoorden.

Psycholinguïstische studies halen meermaals de invloed van het werkwoord aan om de taalverwerving (van kinderen en volwassenen, alsook vreemde taalverwerving) te verklaren en bijgevolg ook om de verwerving van alteminerende woordvolgordes te motiveren (o.a., Holmes, Stowe & Cupples 1989; MacDonald 1994; Bybee & Hopper 2001; Ellis 2002a; Kidd, Lieven & Tomasello 2006). De informatie die taalgebruikers krijgen aan de hand van een werkwoord blijkt daarbij doorslaggevend:

At higher levels, it can be shown that language comprehension is determined by the listeners’ vast amount of statistical information about the behavior of lexical items in their language and that, at least for English, verbs provide some of the strongest constraints on the resolution of syntactic ambiguities. Comprehenders know the relative frequencies with which individual verbs appear in different tenses, in active versus passive structures and in intransitive versus transitive structures, the typical kinds of subjects and objects that a verb takes, and many other such facts. Such information is acquired through experience with input that exhibits these distributional properties. (Ellis 2002a: 144)

Onderzoek toont verder aan dat bij werkwoorden die door twee structuren gevolgd kunnen worden, elk individueel werkwoord een voorkeur heeft voor een van de beide opties (zie ook §3.2.4.3 Collostructional Analysis van Gries & Stefanowitsch 2004; Tily et al. 2008). In het onderzoek van Holmes, Stowe & Cupples (1989); Ellis (2002a) en Kidd, Lieven & Tomasello (2006) gaat het in het bijzonder over werkwoorden die door een bijzin of door een direct object gevolgd kunnen worden (zoals het werkwoord *know*) en in het onderzoek van Tily et al. (2008) over de datiefalternantie; maar ook in onderhavige studie kunnen de werkwoorden gecombineerd worden met een VzC in het middenveld en/of in het achterveld (en dan zeker werkwoorden die een object of complement in de vorm van een VzC oproepen). Een van de selectiecriteria was namelijk dat de VzC's verplaatsbaar moesten zijn (zie §3.1.2.4). De relatieve frequentie waarin een werkwoord met een bepaalde constructie gecombineerd wordt, speelt volgens de eerder vermelde onderzoeken een prominente rol in de structurele keuze die een taalgebruiker maakt.²⁸ Deze redenering past ook binnen de zogenaamde *usage-based approach* van taalverwerving en taaltheorie (cf. Bybee & Hopper 2001; Bybee 2006, Bybee & Beckner 2011; Bybee 2013). We zouden dus kunnen verwachten dat het werkwoord de woordvolgorde bepaalt en bijgevolg ook de positie van VzC's.

²⁸ Onder anderen Holmes et al. (1989) demonstreren hoe deze lexicale voorkeur de leestijd beïnvloedt. Werkwoorden die meestal gecombineerd worden met een bijzin roepen een bijzinsconstructie op nadat taalgebruikers het werkwoord gelezen hebben. Omgekeerd verwachten taalgebruikers bij werkwoorden die vaker gecombineerd worden met een direct object, een objectconstructie. Krijgt de participant tijdens een *self-paced single-word reading task* een intuïzin, dan bemoeilijkt dat de verwerking met een langere leestijd tot gevolg.

Hoofdstuk 4

Resultaten

Often only quantitative methods can separate the wheat from the chaff (Gries 2013: 4-5)

In dit hoofdstuk zullen we stapsgewijs de resultaten van ons corpusgebaseerde onderzoek naar de variabele positie van VzC's in het Nederlands presenteren. In hoofdstuk 2 is het duidelijk geworden dat de variatie over het algemeen verklaard wordt aan de hand van twee principes, meer bepaald het complexiteitsprincipe en het thematische principe, en dat er daarnaast ook nog enkele andere variabelen zijn die een rol lijken te spelen bij de plaatsing van VzC's in de Nederlandse (bij)zin (bv. Jansen 1978, 1979, 2003; Braecke 1990; Hoeksema 2014; voor een overzicht zie tabel 4 en 5). Desondanks is het nog steeds onduidelijk hoe deze principes en variabelen zich precies tot elkaar verhouden en hoeveel syntactische variatie ze kunnen verklaren en voorspellen. In dit hoofdstuk willen we daarom een antwoord bieden op deze en andere losse eindjes die we in de inleiding en in het literatuurstuk opgemerkt hebben. Zo is de invloed van het werkwoord op dit type van syntactische variatie nog niet onderzocht en bijgevolg onbekend, terwijl studies naar andere types van syntactische variatie wel veel aandacht besteden aan het werkwoord en de positie van het werkwoord in de zin (zeker vanuit een cognitieve invalshoek). Kort samengevat willen we met deze studie de volgende vragen beantwoorden:

- Spelen het complexiteitsprincipe (*syntactische complexiteit*) en het thematische principe (*discoursstatus*) allebei een rol, en zo ja, hoe verhouden deze twee verklarende principes zich precies tot elkaar?
- Rekening houdend met de verschillende operationalisaties van de factoren *syntactische complexiteit* en *discoursstatus*, die we in het vorige hoofdstuk gepresenteerd hebben, welke operationalisatie(s) verklaart het best dit type van woordvolgordevariatie?

- Wat is het effect van andere (al dan niet nieuwe) variabelen, zoals de grammaticale functie van de VzC, de betekenis van de VzC, de frequentie van het werkwoord, de betekenis van het werkwoord, het type werkwoord en de zinsrelator?

Bovenstaande vragen zullen we beantwoorden door gebruik te maken van een multivariate statistische techniek die in het verleden al meermaals is ingezet om complexe vormen van syntactische variatie te bestuderen, m.n. logistische regressie. Met die techniek kunnen we heel precies bepalen of en in welke mate de in het vorige hoofdstuk voorgestelde principes en variabele(n) de keuze tussen middenveld- en achterveldpositie kunnen verklaren en voorspellen, rekening houdend met elkaars effect.

Vooraleer we de multivariate regressieanalyse presenteren (§4.2), willen we in een eerste stap de distributie van de variabelen waarvan we vermoeden dat ze een rol kunnen spelen individueel inspecteren. Onder anderen Baayen (2008) en Gries (2013) raden een dergelijke kwantitatieve exploratie van de distributie van de individuele variabelen aan vooraleer ze toe te voegen aan een multivariate analyse (“knowledge of the probability distribution of a [...] variable is often crucial for statistical analysis”, Baayen 2008: 21). Een preliminaire exploratie is dus vooral nuttig omdat het ons meer inzicht geeft in de variabelen en in de verschillende levels binnen eenzelfde variabele. Zo kunnen we simpelweg vaststellen of onze dataset lange middenvelden bevat, of de verhouding definitieve VzC en indefinitieve evenredig is, enz. Zo’n exploratie gebeurt aan de hand van visuele middelen, zoals histogrammen, staafdiagrammen en boxplots, en beschrijvende statistiek, zoals gemiddelde, mediaan, modus, standaardafwijking, interkwartiele range en scheefheid.

In een tweede stap zullen we de relevante variabelen waarvan we vastgesteld hebben dat ze niet correleren (en dus niet-redundante informatie bevatten), toevoegen aan een logistisch regressiemodel. We beschrijven de resultaten van dit multivariate model in paragraaf 4.2 en gaan na wat het precieze effect is van elk van de variabelen en hoeveel van de totale variatie ze kunnen verklaren en voorspellen. De interpretatie van de resultaten en de manier waarop deze studie de huidige kennis van de plaatsing van VzC’s in Nederlandse geschreven bijzinnen uitbreidt, komt in het volgende hoofdstuk aan bod.

Voor alle statistische tests die we in deze paragraaf gebruiken, is het belangrijk op te merken dat het gebruik van deze tests geen doel op zich is, maar wel een middel om betrouwbare uitspraken te doen over de variabelen en over de woordvolgorde van VzC’s. Dat houdt ook in dat we enkel uitspraken kunnen doen over de variabelen zoals wij ze hebben geoperationaliseerd en over de geschreven Belgisch-Nederlands journalistieke taal, die aan alle selectiecriteria (zie §3.1.2) voldoen.

4.1 Een descriptieve exploratie van de verklarende variabelen

In de volgende paragrafen geven we een descriptief overzicht van de afhankelijke variabele, de positie van de VzC (§4.1.1), en van de verschillende onafhankelijke variabelen (van §4.1.2 tot §4.1.5). De vier groepen onafhankelijke variabelen, zoals we ze gepresenteerd hebben in de inleiding en in het vorige methodologische hoofdstuk, worden daarbij achtereenvolgens besproken: *syntactische complexiteit*, *discoursstatus*, *grammaticale en semantische classificatie van de VzC* en *het werkwoord*. We gaan ook na of de verschillende operationalisaties van *syntactische complexiteit* (lengte in woorden, lettergrepen en inbedding) correleren en of de variabelen *definietheid* en *toegankelijkheid*, twee operationalisaties die de bekendheid van de VzC willen vaststellen, gerelateerd zijn.

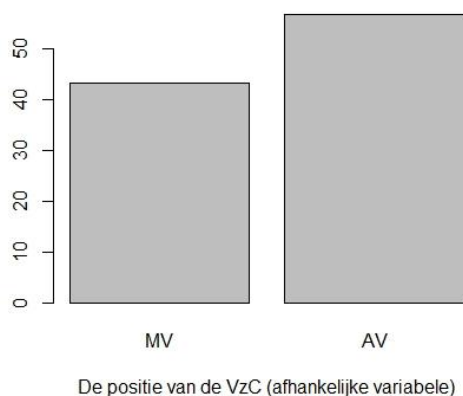
Om de verdeling van de kwantitatieve of numerieke variabelen (bijvoorbeeld de *lengte van het middenveld*, *achterveld* en de *VzC*) te inspecteren, gebruiken we steeds een histogram en een boxplot. Een histogram is waarschijnlijk de meeste gebruikte plot om de frequentieverdeling van een numerieke variabele weer te geven, terwijl de boxplot dan weer de meest gebruikte vorm is om de vijf-getallensamenvatting (minimum, 1^{ste} kwartiel, mediaan, 3^{de} kwartiel en maximum) van een numerieke variabele visueel voor te stellen (Gries 2013). De numerieke variabelen die we in deze studie gebruiken zijn allemaal discreet van aard (de lengte van het middenveld kan niet 2,2 woorden lang zijn) en ze hebben een ratiomeetniveau (we spreken daarom ook over een ratiovariabele), d.w.z. dat er een vast nulpunt is (een waarde 0 geeft de afwezigheid aan van woorden), een vaste afstand tussen de getallen (lengte 4 is wel degelijk dubbel zo lang als lengte 2) en een oplopende volgorde tussen de waarden.

Voor de kwalitatieve of categorische variabelen, zoals *definietheid* en de afhankelijke variabele *positie van de VzC*, gebruiken we een staafdiagram, aangezien “bar plots are probably the most frequent forms of representing the frequencies of nominal/categorical variabls” (Gries 2013: 112). De categorische variabelen in deze studie hebben verschillende meetniveaus. *Definietheid* is een voorbeeld van een nominale variabele: er zijn maar een beperkt aantal categorieën (twee in het geval van *definietheid*) en tussen deze categorieën bestaat er geen inherente volgorde of rangorde. Daardoor is het niet mogelijk om een minimum, maximum of andere descriptieve statistische waarden te berekenen. De variabele *toegankelijkheid* (een van de operationalisaties van de *discoursstatus*) heeft echter wel een rangorde (van heel toegankelijk informatie naar nieuwe informatie), maar ondanks deze rangorde heeft het weinig zin om waarden zoals minimum of maximum te berekenen. Het minimum en maximum zijn namelijk levels en geen waarden zoals bij een numerieke variabele. De variabele *toegankelijkheid* is een voorbeeld van een ordinale variabele.

Voor de overzichtelijkheid zullen we de belangrijkste eigenschappen van elk van de variabelen en de manier waarop ze geoperationaliseerd zijn schematisch aan het begin van elke paragraaf weergeven.

4.1.1 De positie van de VzC

We starten deze descriptieve exploratie met een overzicht van het aantal VzC's in middenveldpositie en in achterveldpositie. Vanuit de traditionele veronderstelling dat het middenveld de topologische standaardpositie is voor VzC's, verwachten we meer VzC's in het middenveld dan in het achterveld. Figuur 11 toont echter aan dat deze verhouding niet overeenstemt met de verwachting: 43% van de VzC staat in het middenveld en 57% in het achterveld. Dit verschil is ook statistisch significant met een compleet evenwichtige verdeling voor middenveld- en achterveldpositie ($X^2 = 15.27$, $df = 1$, $p = 9,33e^{-05}$).



Figuur 11: Staafdiagram van de afhankelijke variabele: overzicht van de relatieve frequentie (y-as) middenveldpositie en achterveldpositie (x-as)

De waargenomen verhouding lijkt ook het eerdere empirische onderzoek van Jansen (1978, 1979) en Braecke (1990) tegen te spreken. Jansen stelde namelijk vast dat slechts 40% van de VzC's met een accent in het achterveld staat, dat in de formele stijl 64,1% van de VzC in het middenveld staat en in de informele stijl zelfs 67,2%. In het onderzoek van Braecke, waarbij geen onderscheid werd gemaakt tussen een formele en informele stijl, stond ongeveer 60% van de VzC's in het middenveld. We voegen hier echter meteen aan toe dat de resultaten van Jansen en Braecke gebaseerd zijn op gesproken data en die in onderhavige studie op geschreven data. Het middenveld kan dus nog steeds de standaardpositie zijn in gesproken Nederlands, maar is dat niet in geschreven Nederlands, toch niet voor geschreven journalistieke teksten. We zullen dit resultaat verder uitwerken in paragraaf 5.1.

4.1.2 De syntactische complexiteit

Operationalisatie 1: *lengte* van het middenveld, achterveld en de VzC

- hoe geoperationaliseerd?
 - (a) aantal woorden;
 - (b) aantal lettergrepen
- Type variabele en meetniveau:
 - discrete numerieke variabelen met ratiomeetniveau

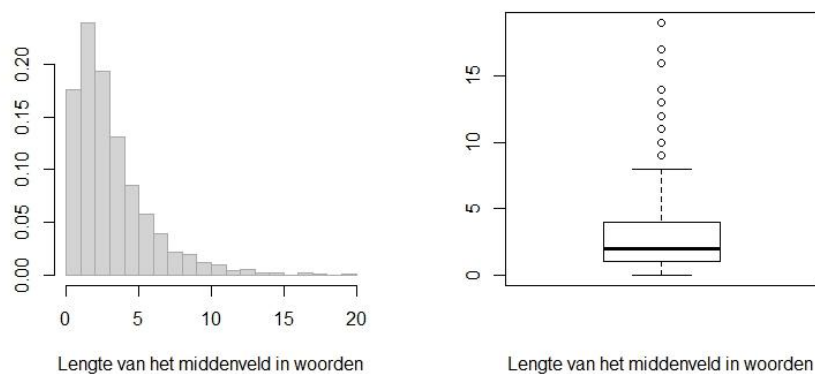
Operationalisatie 2: *inbedding* van het middenveld, achterveld en de VzC

- hoe geoperationaliseerd?
Som van:
 - aantal nabepalingen (geen bijzinnen)
 - aantal bijstelling (geen bijzinnen)
 - aantal bijzinnen als nabepaling of bijstelling*2
- Type variabele en meetniveau:
 - discrete numerieke variabele met ratiomeetniveau

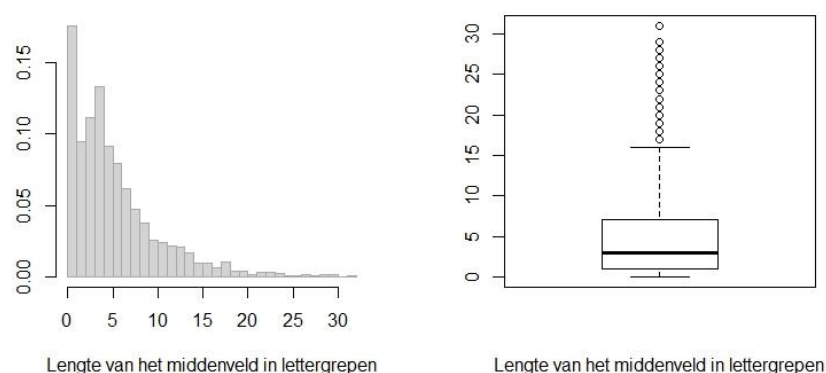
4.1.2.1 Syntactische complexiteit van het middenveld: *lengte* en *inbedding*

De *lengte* van het middenveld *in woorden* en *in lettergrepen* zijn twee discrete numerieke variabele met een ratiomeetniveau. We gebruiken daarom een histogram (links) en een boxplot (rechts). Figuren 12 en 13 geven de verdeling van deze twee variabelen weer.

De tweede operationalisatie, de *inbedding*, wordt in figuur 14 gepresenteerd. De *inbedding* is eveneens een numerieke variabele met een ratiomeetniveau, maar in tegenstelling tot de *lengte* tellen we voor deze laatste variabele het aantal ingebedde structuren en het type *inbedding*. Elk type *inbedding* krijgt een waarde, waarbij de waarde van een bijzin wordt verdubbeld (in navolging van Szmrecsanyi 2004). We beseffen dat deze waarden enigszins arbitrair zijn. Kunnen we bijvoorbeeld beweren dat een bijzin als nabepaling exact dubbel zo complex is als een VzC als nabepaling? Daarom zouden we in dit geval ook over een schaal van *zinnen met geen inbedding* tot *zinnen met meerdere inbeddingen* kunnen spreken. De reden waarom we de variabele toch als een numerieke variabele met een ratiomeetniveau beschouwen, is de aanwezigheid van een absoluut nulpunt. Een variabele met een ordinaal meetniveau houdt ook enkele consequenties in. Bij een ordinale schaal is de oplopende volgorde wel duidelijk, maar zijn de verschillen tussen de waarden moeilijk te interpreteren, waardoor het vreemd is om met deze ordinale waarden berekeningen te maken. Voor de visualisatie van de relatieve frequentie gebruiken we in navolging van de *lengte* een histogram, voor de vijfgetallensamenvatting een boxplot.



Figuur 12: Histogram (links) en boxplot (rechts) van de *lengte in woorden van het middenveld*



Figuur 13: Histogram (links) en boxplot (rechts) van de *lengte in lettergrepen van het middenveld*

De x-assen in histogrammen 12 en 13 representeren de lengte van het middenveld, respectievelijk in woorden en in lettergrepen. In figuur 12 loopt de lengte van het middenveld van 0 woorden (minimum) tot 19 woorden (maximum), en in figuur 13 van 0 lettergrepen (minimum) tot 31 lettergrepen (maximum)¹. De histogrammen zijn verder zo geschaald dat de som van alle staven gelijk is aan 1 (of 100%). Daardoor kunnen we bijvoorbeeld vaststellen dat de observaties met een middenveld van 0 woorden, 1 woord en 2 woorden ongeveer 60% van de dataset representeren (de staaf voor 0 woorden ligt tussen 15% en 20%, de staaf voor 1 woord reikt net geen 25% en de staaf voor 2 woorden net geen 20%).

De histogrammen tonen verder ook aan dat de observaties voor de variabele *lengte* scheef verdeeld zijn. De piek ligt links met een uitloper (of staart) naar rechts. Deze scheefheid (of *skewness*) kunnen we ook berekenen. Een normale verdeling heeft een waarde 0 voor scheefheid. Indien er een rechtse piek is met een uitloper naar links, zal de waarde negatief zijn. Bij een linkse piek met een uitloper naar rechts, wat we verwachten op basis van figuren 12 en 13, is de waarde positief. Conform de verwachting krijgen we

¹ De minimum- en maximumwaarden vinden we ook terug in de boxplot.

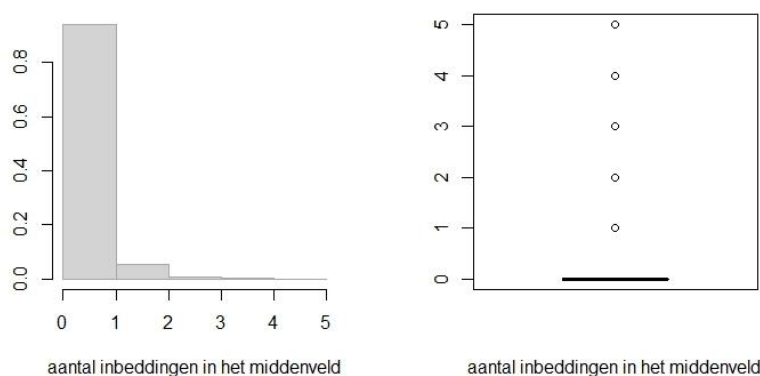
een positieve scheefheidswaarde van 1,80 voor *lengte in woorden* en 1,72 voor *lengte in lettergrepen*.

De boxplots (zie rechts in figuren 12 en 13) tonen eveneens het minimum en maximum, maar voegen extra informatie toe. Zo stellen we in figuur 12 vast dat zinnen met 8 of meer woorden in het middenveld als *outliers* worden beschouwd (de ronde punten). Outliers of extreme waarden beïnvloeden ook het gemiddelde (zie ook Gries 2013: 119 voor een voorbeeld), vandaar dat het belangrijk is om naast het gemiddelde (en de mediaan), ook de standaardafwijking² mee te geven. Voor de *lengte in woorden* is de mediaan 2 (de vette lijn in de boxplot), het gemiddelde is 2,61 en de standaardafwijking is 2,59. Voor de *lengte in lettergrepen* is de mediaan 3, het gemiddelde 4,74 en de standaardafwijking 4,82.

De boxplots tonen verder aan dat het minimum, het 1^{ste} kwartiel en de mediaan dicht bij elkaar liggen in vergelijking met de afstand tussen de mediaan, het 3^{de} kwartiel en het maximum. De boxplots zijn dus niet evenwichtig verdeeld, wat ook wijst op een asymmetrische verdeling rond de mediaan.

Figuur 14 toont aan dat meer dan 80% van de observaties geen inbedding hebben (noch een nabepaling, noch een bijstelling, noch een bijzin). Deze linkse piek zorgt opnieuw voor een scheve verdeling met een positieve waarde 2,82 voor scheefheid.

Observaties met 1 of meer inbeddingen zijn dus veeleer uitzonderlijk en worden daarom in de boxplot als extreme waarden beschouwd. De mediaan is 0, het gemiddelde 0,24 en de standaardafwijking 0,59.



Figuur 14: Histogram (links) en boxplot (rechts) van de *inbedding van het middenveld*

Tot slot berekenen we de Shapiro-Wilk-test, die aangeeft of de observaties normaal verdeeld zijn. Hoewel het ondertussen wel duidelijk is dat dit niet het geval is voor de *syntactische complexiteit* van het middenveld, voegen we er voor de volledigheid de resultaten van de Shapiro-Wilk-test aan toe:

² Hoe groter de spreidingsbreedte (het verschil tussen minimum en maximum) des te groter is de standaarddeviatie.

	Shapiro-Wilk-test	
	waarde	P
Lengte middenveld woorden	0,82	$< 2,2e^{-16}$
Lengte middenveld lettergrepen	0,84	$< 2,2e^{-16}$
Complexiteit middenveld	0,46	$< 2,2e^{-16}$

Tabel 18: Shapiro-Wilk-test voor de verschillende operationalisaties van *syntactische complexiteit* middenveld

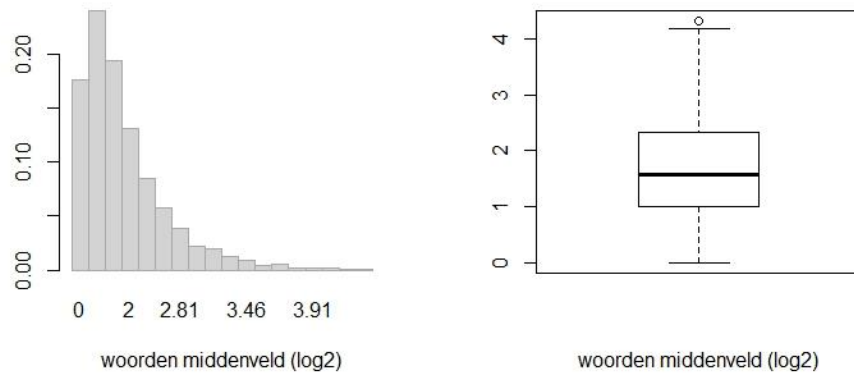
Bij een p-waarde kleiner dan 0,05 kan de nulhypothese, namelijk de hypothese dat de *lengte* (in woorden en in lettergrepen) en het *aantal inbeddingen* normaal verdeeld zijn, verworpen worden.

De exploratieve beschrijving van de syntactische complexiteit van het middenveld illustreert hoe de operationalisaties grote overeenkomsten vertonen. Ze zijn alle drie niet normaal verdeeld, hebben een linkse piek en een uitloper naar rechts (vandaar ook rechtse scheefheid). Het is bekend dat variabelen die erg scheef verdeeld zijn logaritmisch getransformeerd worden, alvorens ze op te nemen in een multivariaat model (o.a. Baayen 2008, Gries 2013). In navolging van Gries houden we daarbij rekening met een binaire logaritmische transformatie (\log_2) i.p.v. \log_{10} :

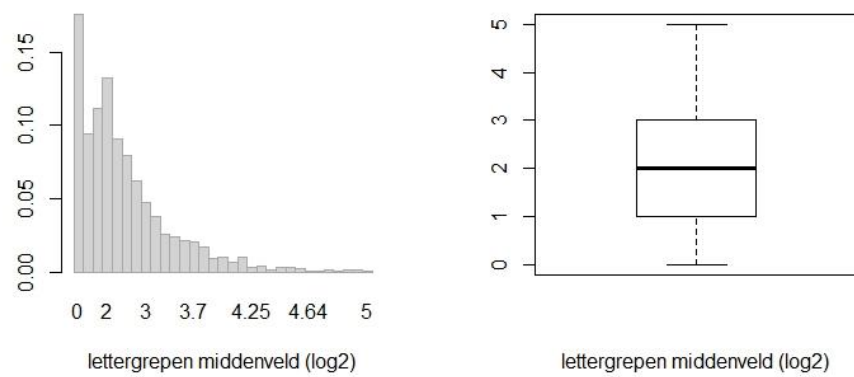
[...] not all numeric predictors should be entered into model as is. For example, frequency effects often operate on a logarithmic scale such that, even if word1 is ten times as frequent as word 2, the effect of word1 on the dependent variable may only be \log_{10} times as strong. Thus, what one should maybe put into the regression equation is $\log(\text{frequency})$ (and to interpret results more easily, it may be good to use logs to the base of 2!). (Gries 2013: 254)

Vervolgens visualiseren we de \log_2 -transformaties voor de *lengte in woorden*, in *lettergrepen* en het *aantal inbeddingen* van het middenveld.³ We gebruiken daarvoor opnieuw een histogram in combinatie met een boxplot.

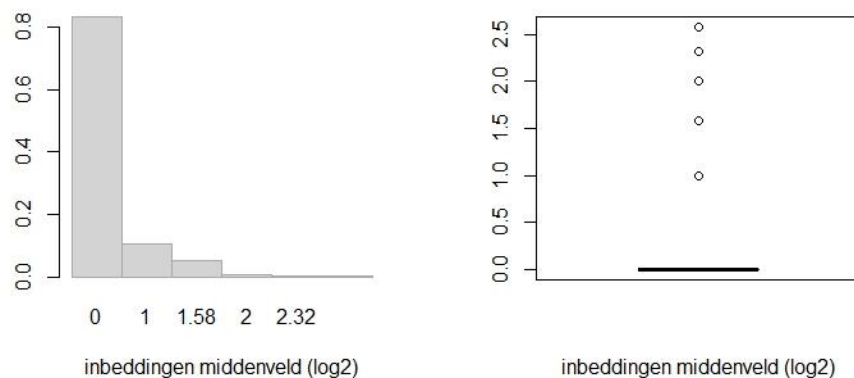
³ Omdat de variabelen heel wat nulwaarden bevatten (0 woorden, 0 lettergrepen en 0 inbeddingen) en de \log_2 van 0 oneindig is, verhogen we elke waarde met 1. Dit ligt in lijn met de $\log_1 p()$ functie in R voor een natuurlijk logaritme.



Figuur 15: Histogram (links) en boxplot (rechts) van de log2-transformatie van het *aantal woorden in het middenveld*



Figuur 16: Histogram (links) en boxplot (rechts) van de log2-transformatie van het *aantal lettergrepen in het middenveld*



Figuur 17: Histogram (links) en boxplot(rechts) van de log2-transformatie van het *aantal inbeddingen in het middenveld*

Per log2-variabele krijgen we een nieuwe maximumwaarde: voor *lengte middenveld woorden* is dat 4,32, voor *lettergrepen* 5 en voor het *aantal inbeddingen* 2,58. In vergelijking met de boxplots in figuren 12 en 13 lijkt de verdeling rond de mediaan meer symmetrisch (en dus 'normaler') in figuren 15 en 16 (respectievelijk de *lengte in woorden* en de *lengte in*

lettergrepen).⁴ Bovendien is er slechts één outlier voor het *aantal woorden in het middenveld* en zijn er geen outliers voor het *aantal lettergrepen in het middenveld*. Het aantal outliers blijft voor de operationalisatie in termen van *inbeddingen* ongewijzigd. De boxplot in figuur 17 toont opnieuw aan dat alle zinnen met geen inbeddingen in het middenveld (dus waarde 0) extreme waarden zijn, net zoals we vastgesteld hadden in figuur 14. Deze extreme waarden kunnen we echter niet verwijderen uit de dataset, want in dat geval zouden we zeer interessante attestaties verwijderen.

Het doel van deze logaritmische transformatie is om de oorspronkelijke scheefheid enigszins te reduceren, of in de woorden van Baayen:

The reason that many of the variables that we study in this book are logarithmically transformed is to eliminate or at least substantially reduce the skewing in their distribution. This reduction is necessary for most statistical techniques discussed in this book to work appropriately. Without the logarithmic transformation, just a few extreme outliers might dominate the outcome, partially or even completely obscuring the main trends characterizing the majority of data points. (Baayen 2008: 31)

Voor de lengte in woorden is de scheefheid gereduceerd van 1,80 tot 0,05, voor de lengte in lettergrepen van 1,72 tot -0,18 en voor inbedding van 2,82 tot 2,16.

4.1.2.2 Syntactische complexiteit van het achterveld: *lengte* en *inbedding*

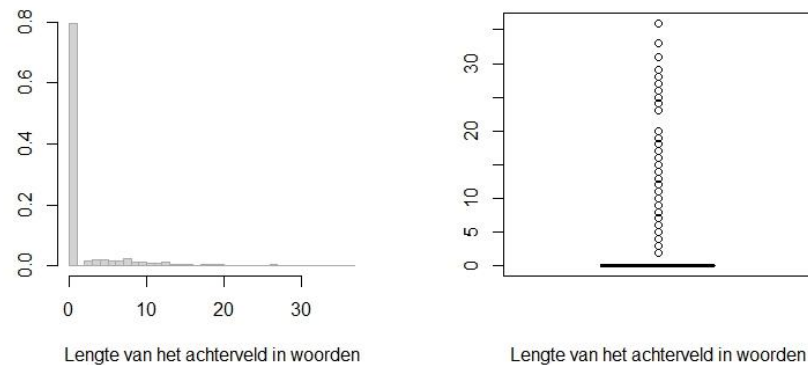
De visuele exploratie van de syntactische complexiteit van het achterveld gebeurt op dezelfde manier als de exploratie van het middenveld. We presenteren een frequentieplot (namelijk een histogram) in combinatie met een boxplot.

De histogrammen voor de *lengte* van het achterveld *in woorden* (links figuur 18) en *in lettergrepen* (links figuur 19) tonen een heel duidelijke linkse piek. Deze piek geeft aan dat de overgrote meerderheid van de zinnen een leeg achterveld hebben. 80% van de observaties hebben namelijk 0 woorden of 0 lettergrepen in het achterveld. Een positieve waarde voor de scheefheid bevestigt deze visuele vaststelling (3,37 voor *lengte in woorden* en 3,46 voor *lengte in lettergrepen*). Verder varieert de lengte in woorden van 0 (minimum) tot 36 (maximum) en in lettergrepen van 0 (minimum) tot 62 (maximum). Deze waarden zijn ook zichtbaar op de x-as van de histogrammen en op de y-as van de boxplots.

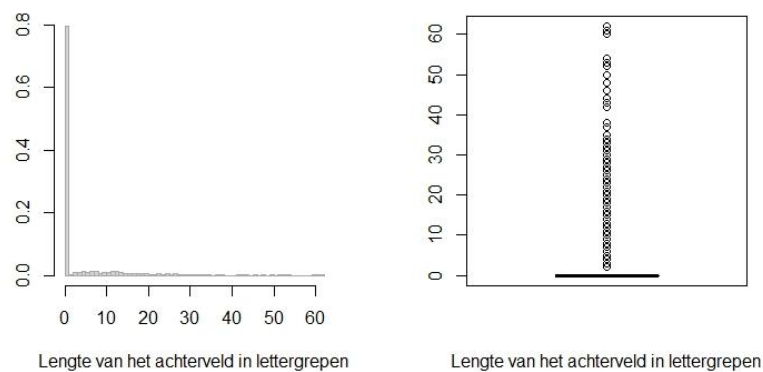
Opvallend in de boxplots (rechts figuren 18 en 19) is dat het 1^{ste} kwartiel, de mediaan en het 3^{de} kwartiel samenvallen. Allemaal hebben ze waarde 0, net zoals de minimumwaarde. Dat betekent ook dat alle observaties (of zinnen) met 1 of meer woorden/lettergrepen in het achterveld als een extreme waarde worden beschouwd.

⁴ Indien we de Shapiro-Wilk-test toepassen op de log2-getransformeerde variabelen, kan de nulhypothese (namelijk dat de data normaal verdeeld zijn) nog steeds verworpen worden.

Deze extreme waarden hebben ook een effect op het gemiddelde, met als resultaat hogere gemiddeldes in vergelijking met de mediaan: 1,84 voor *lengte in woorden* en 3,30 voor *lengte in lettergrepen*. Een grote standaardafwijking (4,69 voor *woorden* en 8,53 voor *lettergrepen*) illustreert dat er een grote spreiding is en dat de waarden ver van het gemiddelde liggen.



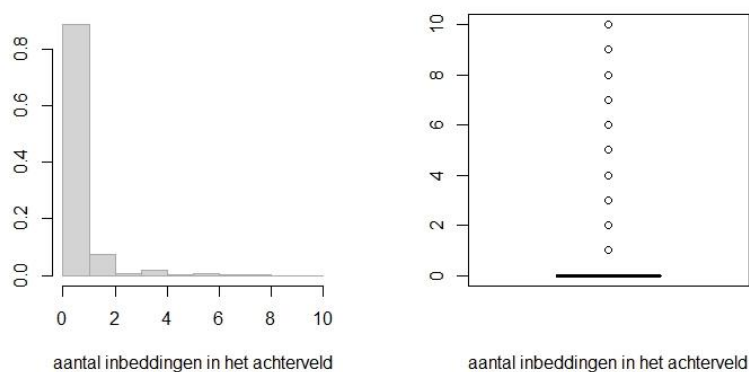
Figuur 18: Histogram (links) en boxplot (rechts) van de *lengte in woorden van het achterveld*



Figuur 19: Histogram (links) en boxplot (rechts) van de *lengte in lettergrepen van het achterveld*

Het histogram in figuur 20 komt zo goed als helemaal overeen met de histogrammen in figuren 18 en 19. De meeste observaties (meer dan 80%) hebben een achterveld met geen *inbeddingen*. De scheefheid die het histogram toont, hebben we opnieuw berekend. Voor de operationalisatie in termen van inbedding is de waarde, zoals verwacht, positief (3,85).

Voor de operationalisatie in termen van *inbedding* is de mediaan 0 (net zoals voor de *lengte in woorden* en *in lettergrepen*), het gemiddelde 0,36 en de standaardafwijking 1,05.



Figuur 20: Histogram (links) en boxplot (rechts) van de *inbedding* van het achterveld

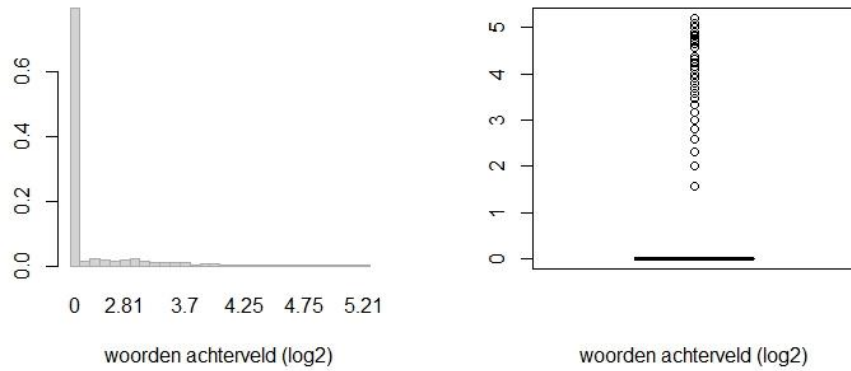
Het aantal inbeddingen varieert van 0 (geen inbedding) tot maximumwaarde 10 (zie ook x-as histogram of y-as boxplot). In dat laatste geval gaat het om één zin, namelijk zin 101. In zin 101 bestaat het achterveld uit 5 ondergeschikte bijzinnen (voorgesteld in tabel 19). Zoals aangehaald in de methodologie wordt de waarde voor een bijzin verdubbeld (5 keer waarde 2 is 10).

- (101) Volgens de ervaren bedrijfsconsultant is het in complexe situaties onmogelijk <begin>dat <su>een manager</su> <vz>over de nodige objectiviteit</vz> <v>beschikt</v> om altijd te weten wat te doen en zeker te zijn dat de mensen die aan hem rapporteren op dezelfde eenvormige manier handelen in hun respectieve businessunits of diensten <end>

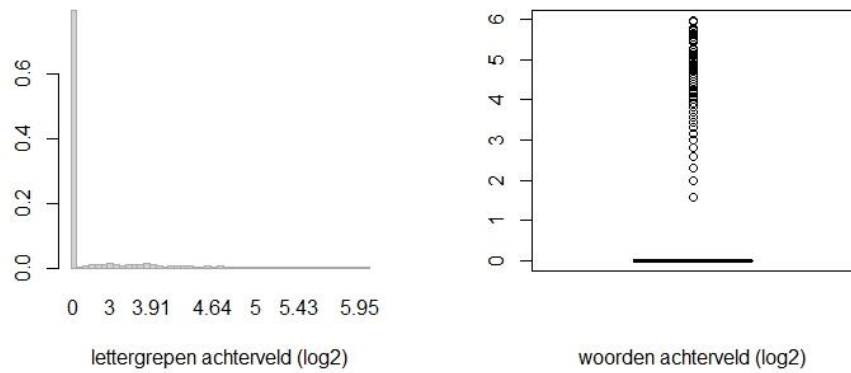
Niveau		Bijzintype
1	om altijd te weten	Beknopte bijzin
2	wat te doen	Beknopte bijzin
#	En	
1	zeker te zijn	Beknopte bijzin
2	dat de mensen	Grammaticale voegwoordzin
3	die aan hem rapporteren	Betrekkelijke bijzin
2'	op dezelfde eenvormige manier handelen in hun respectieve businessunits of diensten	

Tabel 19: Zin 101 ontleed; hiërarchie (of de niveaus) in kolom 1 en het bijzintype in kolom 3

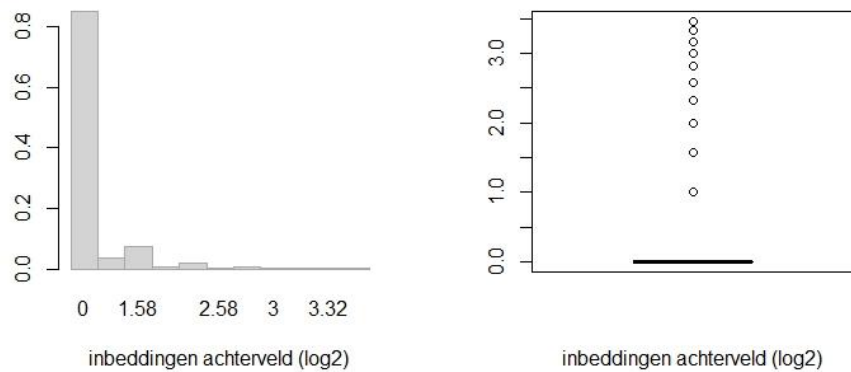
Samengevat zijn de drie operationalisaties voor de syntactische complexiteit van het achterveld vergelijkbaar en vertonen ze grote overeenkomsten (min of meer dezelfde frequenties en een duidelijke linkse piek of rechtse scheefheid). Omdat deze variabelen niet normaal verdeeld zijn (Shapiro-Wilk-test: $W = 0,46$ voor *lengte in woorden* ($p < 2.2e^{-16}$), $W = 0,45$ voor *lengte in lettergrepen* ($p < 2.2e^{-16}$) en $W = 0,40$ voor *aantal inbeddingen* ($p < 2.2e^{-16}$)), zullen we de variabelen logaritmisches transformeren (\log_2). De verdeling van deze getransformeerde variabelen worden in onderstaande figuren, respectievelijk *woorden*, *lettergrepen* en *inbeddingen*, gepresenteerd.



Figuur 21: Histogram (links) en boxplot (rechts) van de \log_2 -transformatie van het *aantal woorden in het achterveld*



Figuur 22: Histogram (links) en boxplot (rechts) van de \log_2 -transformatie van het *aantal lettergrepen in het achterveld*



Figuur 23: Histogram (links) en boxplot (rechts) van de \log_2 -transformatie van het *aantal inbeddingen in het achterveld*

De boxplots in figuren 21, 22 en 23 zijn op het eerste gezicht vergelijkbaar met de boxplots in figuren 18, 19 en 20. Toch zijn de waarden geschaald, is de spreiding minder groot (van 0 tot maximum 5,21 voor *woorden*, maximum 5,98 voor *lettergrepen* en tot maximum 3,46 voor *inbeddingen*) en is de oorspronkelijke scheefheid gereduceerd (van

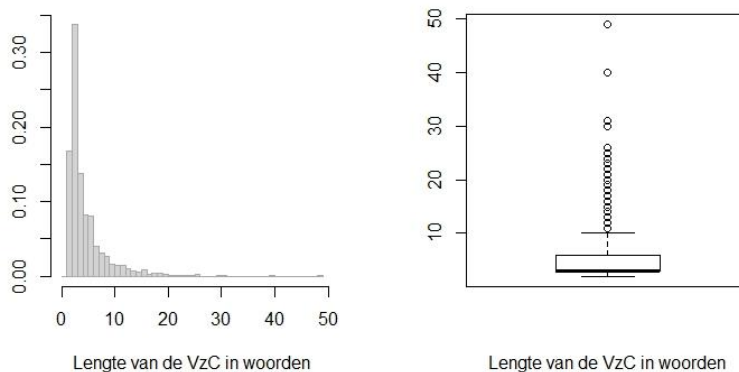
3,37 tot 1,80 voor *woorden*, van 3,46 tot 1,74 voor *lettergrepen* en van 3,84 tot 2,46 voor *inbeddingen*).

4.1.2.3 Syntactische complexiteit van de VzC: *lengte* en *inbedding*

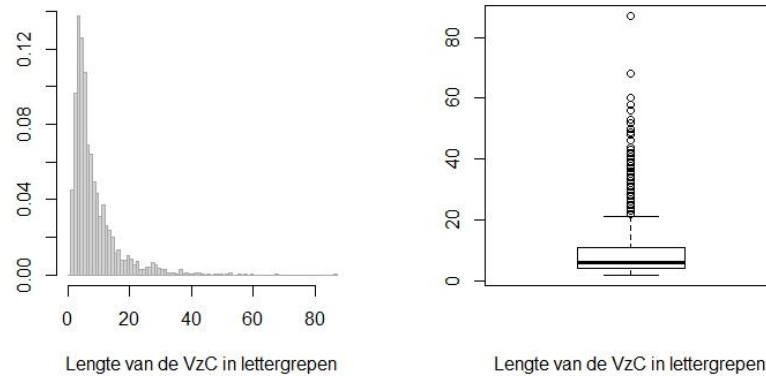
Voor de *lengte* van de VzC in *woorden* en *lettergrepen* zijn er geen observaties met minder dan 2 woorden of lettergrepen. Dat is natuurlijk vanzelfsprekend aangezien een VzC altijd uit minstens 2 woorden moet bestaan (namelijk een voorzetsel gevolgd door een relatum). Net zoals bij de *lengte van het middenveld* representeren de eerste lengtematen het grootste deel van de dataset. VzC's bestaande uit 2 en 3 woorden of uit 2 tot 6 lettergrepen stellen ongeveer 50% van de dataset voor.

Op basis van de histogrammen in figuur 24 en 25 veronderstellen we een scheve verdeling met een rechtse scheefheid (of *skewness*). Om deze waarneming te staven, berekenen we opnieuw de scheefheid voor de *lengte in woorden* en in *lettergrepen*. Een positieve waarde van 3,40 voor *woorden* en 3,01 voor *lettergrepen* wijst op een linkse piek en een rechtse uitloper.

De boxplots (rechts figuren 24 en 25) visualiseren de mediaan, het 1^{ste} en 3^{de} kwartiel, minimum en maximum. Voor de operationalisatie in termen van *woorden* is de mediaan 3, het gemiddelde 4,99 en de standaardafwijking 4,05. Voor de operationalisatie in termen van *lettergrepen* is de mediaan 6, het gemiddelde 9,02 en de standaardafwijking 7,9.



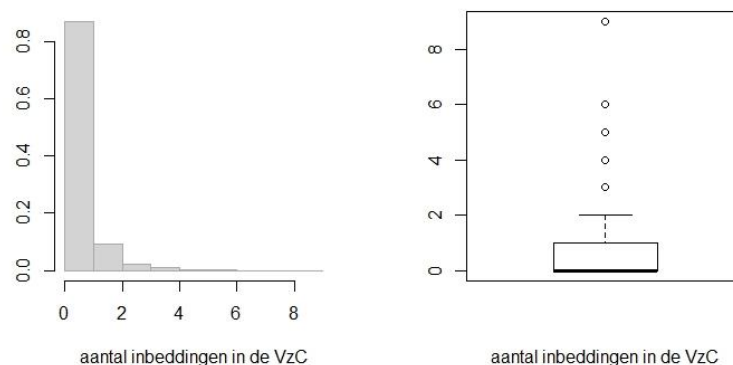
Figuur 24: Histogram (links) en boxplot (rechts) van de *lengte in woorden* van de VzC



Figuur 25: Histogram (links) en boxplot (rechts) van de *lengte in lettergrepen* van de VzC

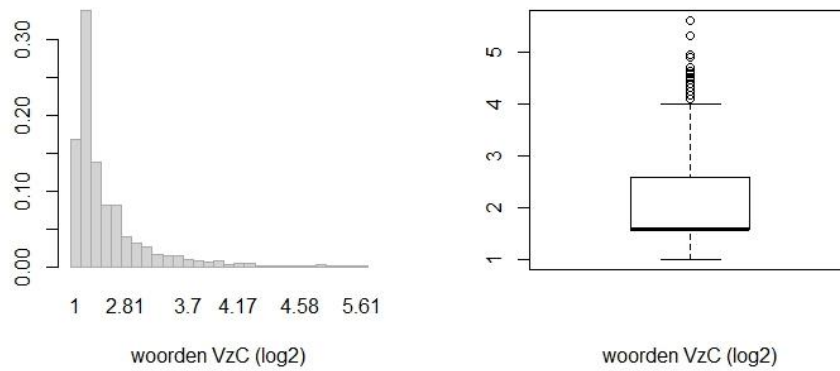
Voor de *inbedding* van de VzC geldt min of meer hetzelfde verhaal als voor de *inbedding* van het *middenveld* en *achterveld*. De meeste zinnen hebben geen inbedding en er is een duidelijke asymmetrische verdeling met een linkse piek en rechtse uitloper (waarde voor scheefheid is 2,70).

In tegenstelling tot de *inbedding* van het *middenveld* en het *achterveld* is de boxplot voor de *inbedding* van de VzC een box en geen lijn (in het geval het 1^{ste} kwartiel, de mediaan en het 3^{de} kwartiel samenvallen). Concreet betekent dat dat VzC's met geen inbedding, met 1 inbedding en 2 inbeddingen niet als extreme waarden worden beschouwd. Dat is wel het geval voor zinnen met een VzC met een waarde 3, 4, 5, 6 en 9 voor inbedding (in totaal stellen ze 4% van de dataset voor).

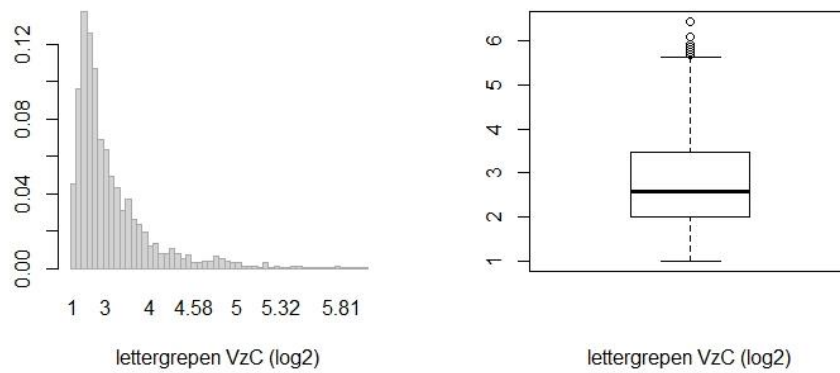


Figuur 26: Histogram (links) en boxplot (rechts) van het aantal *inbeddingen* van de VzC

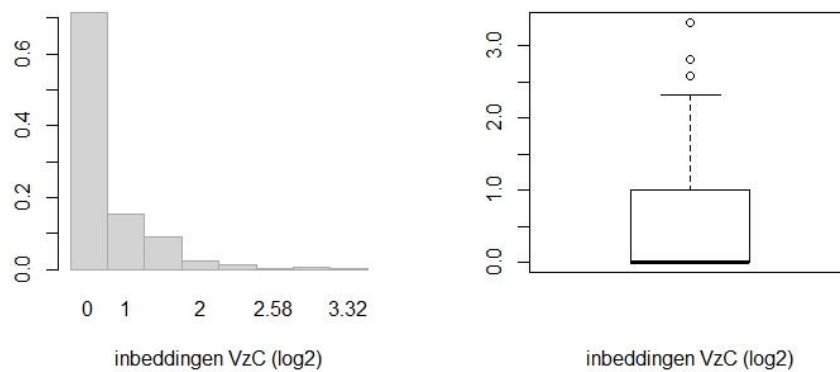
Net zoals de verschillende operationalisaties van de syntactische complexiteit van het middenveld en achterveld, vertonen ook de operationalisaties van de VzC gelijkenissen: de variabelen zijn niet normaal verdeeld (Shapiro-Wilk test: $W = 0,66$ voor *lengte in woorden* ($p < 2.2e^{-16}$), $W = 0,71$ voor *lengte in lettergrepen* ($p < 2.2e^{-16}$) en $W = 0,57$ voor *aantal inbeddingen* ($p < 2.2e^{-16}$)) en ze hebben een uitgesproken rechtse scheefheid. Indien we alle operationalisaties meenemen in een multivariaat model, zullen we deze variabelen eerst logaritmisches transformeren (\log_2). Figuren 27, 28 en 29 visualiseren deze transformaties:



Figuur 27: Histogram (links) en boxplot (rechts) van de log2-transformatie van het *aantal woorden in de VzC*



Figuur 28: Histogram (links) en boxplot (rechts) van de log2-transformatie van het *aantal lettergrepen in de VzC*



Figuur 29: Histogram (links) en boxplot (rechts) van de log2-transformatie van het *aantal inbeddingen in de VzC*

De boxplots in figuren 27, 28 en 29 geven aan dat de waarden geschaald zijn met overeenkomstig nieuwe maxima. Het maximum voor het *aantal woorden* is 4,32, voor het *aantal lettergrepen* 5 en voor het *aantal inbeddingen* 2,58. Net zoals bij de *syntactische complexiteit van het middenveld* en het *achterveld* reduceren de logaritmische transformaties

de scheefheid (van 3,40 tot 0,94 voor *woorden*, van 3,01 tot 0,48 voor *lettergrepen* en van 2,70 tot 1,47 voor *inbedding*).

4.1.2.4 De relatie tussen de complexiteitsoperationalisaties: correlatietest en Principal Component Analysis

In general, almost any data matrix can be simplified by PCA. A large table of numbers is one of the more difficult things for the human mind to comprehend. (Wold et al. 1987: 39)

De syntactische complexiteit van het middenveld, het achterveld en de VzC hebben we op verschillende manieren geoperationaliseerd: *in woorden*, *in lettergrepen* en *in mate van inbedding*. Omdat de operationalisaties in de bovenstaande exploratie zeer sterke gelijkenissen vertonen, gebruiken we een correlatietest en een *Principal Component Analysis* (PCA) om na te gaan of de verschillende operationalisaties van het middenveld, achterveld en de VzC inderdaad correleren. Ook puur intuïtief zouden we kunnen verwachten dat de variabelen correleren: hoe meer woorden, des te meer lettergrepen of hoe meer inbeddingen, des te langer, enz. Variabelen die correleren verklaren (min of meer) dezelfde variatie en kunnen bijgevolg voor multicollineariteit zorgen in een multivariate analyse. Het effect van de onafhankelijke correlerende variabelen op een afhankelijke variabele wordt dan arbitrair – de onafhankelijke variabelen overlappen ten minste gedeeltelijk met elkaar – wat ook de betrouwbaarheid van het model reduceert. Bij multicollineariteit in een multivariaat model geeft men de raad om één van de twee onderling correlerende onafhankelijke variabelen uit de analyse te verwijderen. In onderhavige studie werken we voor de verschillende operationalisaties van syntactische complexiteit omgekeerd. We testen eerst of de variabelen correleren, net omdat de visuele en descriptieve exploratie in die richting wijst en we ook puur intuïtief een vermoeden hebben, en voegen dan pas de variabelen toe aan een multivariate analyse.

We gebruiken twee tests: enerzijds een correlatietest en anderzijds een *principal component analysis* (PCA). De laatste techniek, een PCA, wordt vooral gebruikt om grote datasets te reduceren en om zo zicht te krijgen op de onderliggende relaties tussen de verschillende variabelen. Meer concreet wordt er met behulp van een PCA nagegaan of bepaalde variabelen gecombineerd kunnen worden waardoor de dataset op een ‘simpelere’ manier beschreven kan worden. Deze nieuwe variabelen zijn niet gecorreleerd en worden de *principal components* (PCs) of hoofdcomponenten genoemd. Ze verklaren elk een deel van de variabiliteit in de dataset. PCA is dus een mathematische procedure die het mogelijk maakt om een aantal (mogelijk) gecorreleerde variabelen te transformeren in een kleinere set van niet gecorreleerde hoofdcomponenten. Belangrijk om te weten is dat een PCA niet zal werken indien de relatie tussen de variabelen niet sterk genoeg is. Vandaar dat we, voor we een PCA toepassen, eerst de correlatiecoëfficiënt

berekenen (aan de hand van een correlatietest). Indien de correlatiecoëfficiënt kleiner is dan 0,03, dan zal een PCA niet helpen om de dataset te reduceren.

In de volgende paragraaf 4.1.2.4.1 gaan we eerst na of de *lengte in woorden* en de *lengte in lettergrepen* (voor het middenveld, achterveld en de VzC) correleren. In paragraaf 4.1.2.4.2 bestuderen we de relatie tussen *lengte* en *inbedding*.

4.1.2.4.1 De relatie tussen lengte in woorden en lengte in lettergrepen

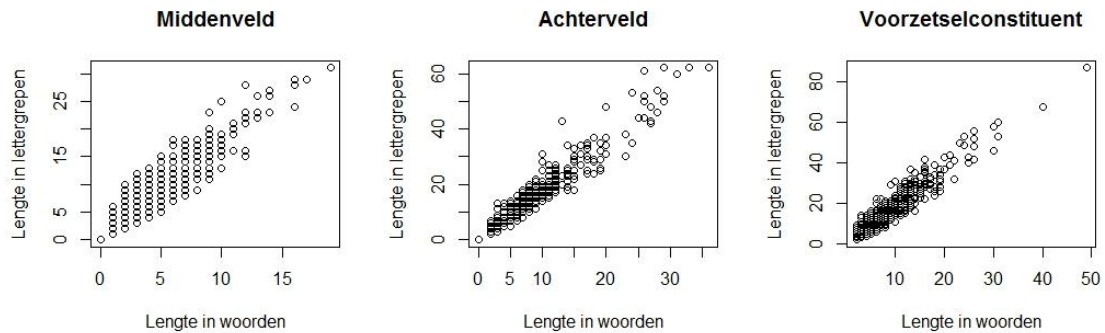
Een correlatietest meet op welke manier twee variabelen samen veranderen. Het resultaat, met name de correlatiecoëfficiënt, beschrijft zowel de sterkte als de richting van deze relatie. Bij ratiovariabelen zoals de *lengte in woorden* en de *lengte in lettergrepen* gebruikt men daarvoor de Pearsons product-momentcorrelatie. De Pearsons correlatie gaat na of er een lineaire relatie bestaat tussen twee variabelen. De correlatiecoëfficiënt kan variëren tussen -1 en +1. Een waarde 0 betekent geen lineaire samenhang, +1 een perfecte positieve lineaire samenhang en -1 een perfecte negatieve lineaire samenhang. De regel is: hoe verder de correlatiecoëfficiënt verwijderd is van 0, hoe sterker de correlatie. Figuur 30 (uit Gries 2013: 147) geeft een meer gedetailleerd overzicht van hoe we de correlatiecoëfficiënt moeten interpreteren.

Correlation coefficient	Labeling the correlation	Kind of correlation
$0.7 < r \leq 1$	very high	positive correlation: the more/higher ..., the more/higher ... the less/lower ..., the less/lower ...
$0.5 < r \leq 0.7$	high	
$0.2 < r \leq 0.5$	intermediate	
$0 < r \leq 0.2$	low	
$r \approx 0$	no statistical correlation (H_0)	
$0 > r \geq -0.2$	low	negative correlation: the more/higher ..., the less/lower ... the less/lower ..., the more/higher ...
$-0.2 > r \geq -0.5$	intermediate	
$-0.5 > r \geq -0.7$	high	
$-0.7 > r \geq -1$	very high	

Figuur 30: Overzicht en interpretatie van de correlatiecoëfficiënt (Gries 2013: 147)

De Pearsons correlatiecoëfficiënt voor het *middenveld in woorden* en *lettergrepen* is 0,941 ($p < 2,2e^{-16}$), voor het *achterveld in woorden* en *lettergrepen* 0,982 ($p < 2,2e^{-16}$) en voor de *VzC in woorden* en *lettergrepen* 0,948 ($p < 2,2e^{-16}$). In de drie gevallen gaat het om een zeer sterke positieve lineaire relaties Deze relaties kunnen we ook visueel voorstellen in de volgende spreidingsdiagrammen⁵:

⁵ We hadden ook omgekeerd kunnen werken: eerst een spreidingsdiagram plotten en op basis daarvan het type correlatietest bepalen. Een Pearson correlatietest beperkt zich namelijk tot een lineair verband terwijl er natuurlijk ook een niet-lineair verband kan bestaan tussen de variabelen.



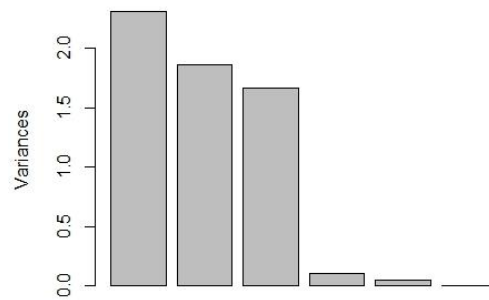
Figuur 31: Grafische voorstelling van de correlerende operationalisaties van *lengte* (in woorden en lettergrepen) van het *middenveld*, *achterveld* en de *VzC*

Op basis van de correlatiecoëfficiënten (en de spreidingsdiagrammen) zouden we kunnen besluiten dat de twee lengte-operationalisaties positief correleren en dat we het verdere onderzoek tot één operationalisatie kunnen beperken.

In een volgende stap voeren we een PCA uit. In navolging van Venables & Ripley (2002) werken we met de getransformeerde logaritmische variabelen.⁶

Met behulp van een PCA kunnen we de zes variabelen (*lengte in woorden* (\log_2) van het *middenveld*, *achterveld* en de *VzC* en *lengte in lettergrepen* (\log_2) van het *middenveld*, *achterveld* en de *VzC*) samen verwerken in een plot en eventuele (onverwachte) relaties tussen de variabelen zien (bijvoorbeeld een relatie tussen *middenveld* en *achterveld*). We berekenen eerst de standaardafwijking en de richting van elk van de zes *principal components* (of hoofdcomponenten of PCs). De resultaten geven we grafisch weer in een staafdiagram (figuur 32): de staven tonen de verklaarde variabiliteit in de dataset (y-as) per PC (x-as). Op basis van dit staafdiagram kunnen we besluiten hoeveel PCs nuttig en nodig zijn in een verdere analyse. In het geval van de twee verschillende lengte-operationalisaties van het *middenveld*, *achterveld* en de *VzC* verklaren de eerste drie PCs de meeste variabiliteit in de dataset. We kunnen het aantal variabelen dus halveren (van 6 variabelen naar 3 hoofdcomponenten).

⁶ Een andere mogelijkheid is om met een Box Cox transformatie te werken.



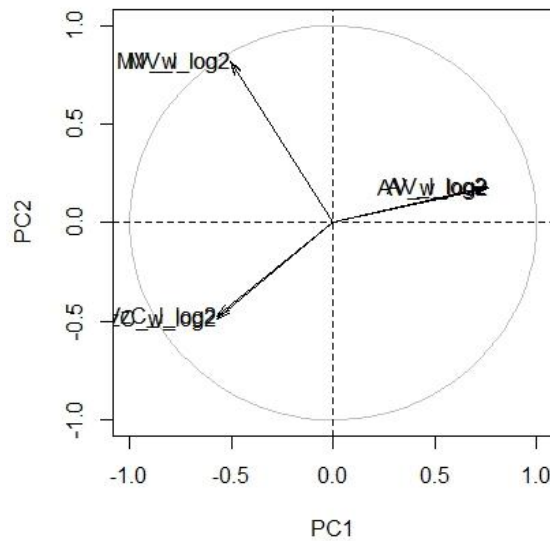
Figuur 32: De relatie tussen *lengte in woorden* en *lengte in lettergrepen*: de relatieve variantie (y-as) per PC (x-as)

We kunnen ook de invloed of het belang van deze PCs berekenen. In tabel 20 toont rij 1 de standaardafwijking van elke PC, rij 2 de verklaarde variantie per PC (proportioneel) en rij 3 de cumulatieve som van de verklaarde variantie (proportioneel). Net zoals in figuur 32 stellen we vast dat de eerste drie PCs zo goed als alle variantie verklaren (meer dan 97%) en dat de laatste drie PCs overbodig zijn in verdere analyses.

Belang van de componenten						
	PC1	PC2	PC3	PC4	PC5	PC6
Standaardafwijking	1.520	1.366	1.290	0.322	0.224	0.070
Variantie (proportioneel)	0.385	0.311	0.278	0.017	0.008	0.001
Cumulatieve variantie	0.385	0.696	0.974	0.991	0.999	1.000

Tabel 20: De relatie tussen *lengte in woorden* en *lengte in lettergrepen*: overzicht van de standaardafwijking, de variantie en de cumulatieve variantie van de 6 hoofdcomponenten

In een laatste stap projecteren we de data en de oorspronkelijke variabelen (*lengte in woorden* (\log_2) en *lengte in lettergrepen* (\log_2) van het *middenveld*, *achterveld* en de *VzC*) op basis van de eerste twee PCs (zij verklaren samen 0.696 % van de variantie). We gebruiken daarvoor een correlatiecirkel (zie figuur 33), waarbij de eerste hoofdcomponent (of PC1) op de x-as geprojecteerd wordt en de tweede hoofdcomponent (of PC2) op de y-as. De pijlen *lengte in woorden* (\log_2) en *lengte in lettergrepen* (\log_2) van respectievelijk het *middenveld* (*MV_w_log2* en *MV_l_log2* in figuur 33), *achterveld* (*AV_w_log2* en *AV_l_log2* in figuur 33) en de *VzC* (*VzC_w_log2* en *VzC_l_log2* in figuur 33) overlappen zo goed als volledig. Dat betekent dat ze correleren. Daarnaast zien we hoe de pijlen voor het *middenveld*, *achterveld* en de *VzC* in een ander kwartiel liggen. Dat betekent dat er geen correlatie is tussen het *middenveld*, *achterveld* of de *VzC*. Samengevat kunnen we op basis van deze plot besluiten dat we slechts één lengte-operationalisatie meenemen in een verdere analyse (óf woorden óf lettergrepen) van zowel het *middenveld*, het *achterveld* én de *VzC*.

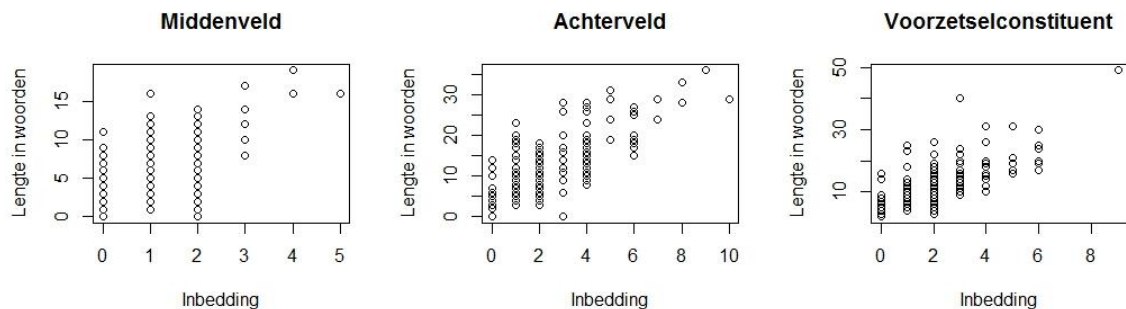


Figuur 33: Correlatiecirkel van de *lengte in woorden* (\log_2) en de *lengte in lettergrepen* (\log_2) op basis van PC1 (x-as) en PC2 (y-as)

4.1.2.4.2 De relatie tussen lengte in woorden en inbedding

In deze paragraaf gaan we na of de *lengte*, als een operationalisatie van syntactische complexiteit, correleert met de andere operationalisatie van syntactische complexiteit, namelijk *inbedding*. In de vorige paragraaf hebben we vastgesteld dat het aantal woorden en het aantal lettergrepen, als operationalisaties van *lengte*, correleren. Daarom maken we in deze paragraaf de vergelijking tussen *inbedding* en *lengte in woorden* (en dus niet tussen *inbedding* en *lengte in lettergrepen*). In wat volgt, presenteren we eerst de correlatietest. Vervolgens voeren we een *principal component analysis* uit.

Om de correlatietest te berekenen, gebruiken we opnieuw Pearsons correlatietest. De Pearsons correlatiecoëfficiënten voor de *lengte* en de *inbedding* van het middenveld is 0,74 ($p < 2,2e^{-16}$), voor het achterveld 0,89 ($p < 2,2e^{-16}$) en voor de VzC 0,85 ($p < 2,2e^{-16}$). Alle waarden liggen tussen 0,7 en 1. Volgens Gries (2013; zie ook figuur 30) bestaat er een zeer sterke positieve lineaire relatie tussen deze operationalisaties. Deze positief lineaire correlaties stellen we ook visueel voor:

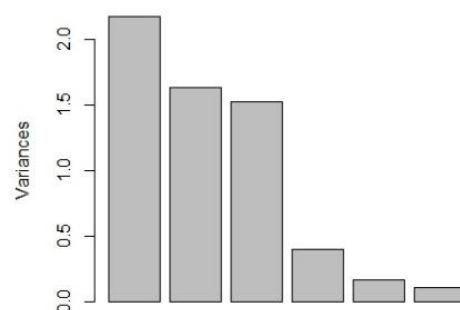


Figuur 34: Grafische voorstelling van *inbedding* en *lengte in woorden* van het middenveld, achterveld en de VzC

Voor het achterveld en de voorzetselconstituent zien we in figuur 34 een stijgende lineaire relatie tussen de *inbedding* en de *lengte in woorden*. Voor het middenveld is deze samenhang minder nadrukkelijk aanwezig, zoals ook al bleek uit de correlatietest; niettemin is er ook voor het middenveld een duidelijke correlatie.

De Pearsons correlatietest veronderstelt een lineaire relatie tussen ratiovariabelen. De variabele *inbedding* konden we echter ook als een ordinale variabele beschouwen, bestaande uit een schaal van niet-complex (want geen inbeddingen) tot zeer complex (voor het middenveld gaat deze schaal tot 5 inbeddingen). Indien een van de variabelen ordinaal is, wordt in de meeste gevallen een Spearman correlatietest⁷ gebruikt in plaats van een Pearson correlatietest. Voor de volledigheid geven we ook de Spearman correlatiecoëfficiënten mee: voor de operationalisaties (*inbedding* en *lengte*) van het middenveld is de coëfficiënt 0,58 ($p < 2,2e^{-16}$); voor het achterveld 0,87 ($p < 2,2e^{-16}$) en voor de VzC 0,76 ($p < 2,2e^{-16}$). Voor het achterveld en de VzC krijgen we opnieuw een zeer sterke monotone relatie tussen de beide operationalisaties. Voor het middenveld is deze relatie sterk (tussen 0,5 en 0,7) zonder meer. Op basis van de beide correlatiecoëfficiënten kunnen we besluiten dat de operationalisatie in termen van *lengte* en *inbedding* correleren.

Vervolgens voeren we een PCA uit. In navolging van Venables & Ripley (2002) werken we opnieuw met de getransformeerde logaritmische variabelen. Figuur 35 stelt vast dat de meeste variabiliteit in de dataset door de eerste drie PCs verklaard wordt. De set van 6 variabelen kan dus gereduceerd worden tot 3 niet-gecorrleerde hoofdcomponenten.



Figuur 35: De relatie tussen *lengte in woorden* en *inbedding*: de relatieve variantie (y-as) per PC (x-as)

Tabel 21 toont de standaardafwijking van elke PC, de proportie verklaarde variantie en de cumulatieve proportie verklaarde variantie. De eerste drie PCs zijn in staat om 88,7% van de variabiliteit in de dataset te vatten.

⁷ De Spearman correlatietest gaat na of er een monotone relatie is tussen de variabelen (in tegenstelling tot Pearsons correlatietest waarbij enkel gezocht wordt naar een lineair verband).

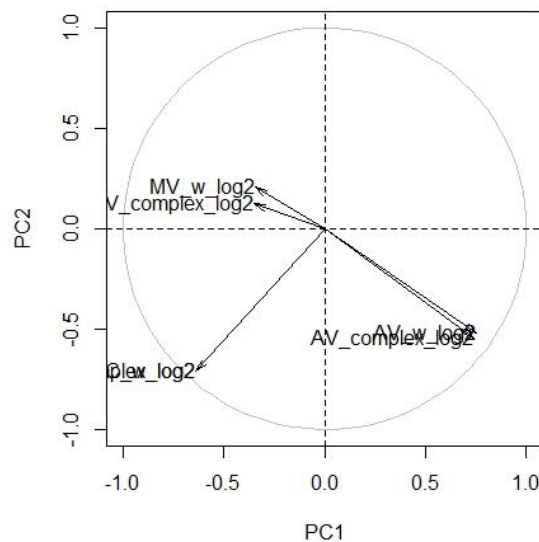
Belang van de componenten						
	PC1	PC2	PC3	PC4	PC5	PC6
Standaardafwijking	1.474	1.277	1.233	0.633	0.411	0.329
Variantie (proportioneel)	0.362	0.272	0.253	0.067	0.028	0.018
Cumulatieve variantie	0.362	0.634	0.887	0.954	0.982	1.000

Tabel 21: De relatie tussen *lengte in woorden* en *inbedding*: overzicht van de standaardafwijking, de variantie en de cumulatieve variantie van de 6 hoofdcomponenten

Net zoals voor de operationalisatie in termen van woorden en lettergrepen visualiseren we de *principal component analysis* met behulp van een correlatiecirkel (met PC1 op de x-as en PC2 op de y-as).

De pijlen voor de lengte en de inbedding van het achterveld (kwartiel rechts onder) en de VzC (kwartiel links onder) overlappen volledig. Deze operationalisaties correleren en we zullen in het vervolg van deze studie opteren om met één operationalisatie te werken. De pijlen voor de operationalisaties van het middenveld (kwartiel links boven) overlappen niet maar de hoek tussen de pijlen is klein genoeg om te zeggen dat ze (min of meer) dezelfde variabiliteit verklaren. Ook in het geval van het middenveld kunnen we dus het best één operationalisatie selecteren.

De pijlen voor middenveld, achterveld en de VzC liggen opnieuw in verschillende kwartielen. Er zijn dus geen opvallende (cor)relaties tussen de velden en/of de VzC vast te stellen.



Figuur 36: Correlatiecirkel van de *lengte in woorden (log2)* en het *aantal inbeddingen (log2)* op basis van PC1 (x-as) en PC2 (y-as)

4.1.2.4.3 Samenvatting van de relatie tussen de complexiteitsoperationalisaties

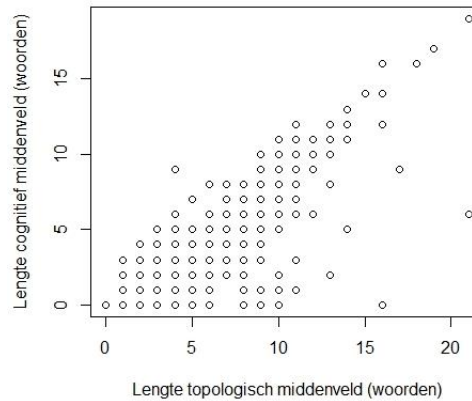
In de methodologie gaven we aan dat we niet a priori konden veronderstellen dat een langere constituent automatisch meer inbeddingen heeft (zie ook Jansen 1978; Beaman 1984; Givón 1991; Szmrecsanyi 2004). Bovendien kon ook de lengte in woorden en de lengte in lettergrepen verschillen (zie Gries 2013 en §3.2.1.1). Dat was ook de reden waarom we met meerdere operationalisaties rekening hebben gehouden. De correlatietests en de *principal component analyses* geven ons echter voldoende bewijs om op dit punt te concluderen dat de verschillende operationalisaties overlappen en dat ze positief correleren. We kunnen bijgevolg besluiten dat één operationalisatie voor het middenveld, achterveld en de VzC voldoende is.

De 9 oorspronkelijke complexiteitsvariabelen worden dus gereduceerd tot 3 variabelen. Omdat de operationalisatie in termen van *woorden* de meest gebruikte is in empirisch onderzoek naar woordvolgordevariatie (Hawkins 1990; Wasow 1997, 2002; Arnold et al. 2000; Szmrecsanyi 2004), zullen wij in het vervolg van deze studie met *lengte in woorden* werken (1. *Lengte van het middenveld in woorden*, 2. *Lengte van het achterveld in woorden*, 3. *Lengte van de VzC in woorden*). Onder andere Wasow (1997) en Shih (2010, 2011) hebben aangetoond dat de operationalisatie in termen van *woorden* een representatieve manier is om syntactische complexiteit te bestuderen.

4.1.2.5 De relatie tussen het topologische middenveld en het cognitief congruerende middenveld

Tot slot gaan we ook dieper in op de verschillende operationalisaties van het zogenaamde middenveld (of van de zogenaamde tangconstructie, zie ook §2.1.3 en §3.2.1.3). Het middenveld in een Nederlandse bijzin kan namelijk gezien worden als die positie die zich tussen het voegwoord en de werkwoordelijke eindgroep bevindt (d.i. de grammaticale topologische tangconstructie: Van de Velde 1973; ANS 1997;) of als die positie die zich tussen het subject van de bijzin en de werkwoordelijke eindgroep bevindt (d.i. de cognitief congruerende tangconstructie: Uhlenbeck 1962; Jansen 1978; Vrisekoop 2001; Bouman 2009). Om deze twee tangconstructies te vergelijken, hebben we het aantal woorden geteld van de topologische tang en het aantal woorden van de cognitieve tang (zie ook §3.2.1.3). Aan de hand van een correlatietest en een *principal component analysis* kunnen we nagaan wat de relatie tussen de beide precies is.

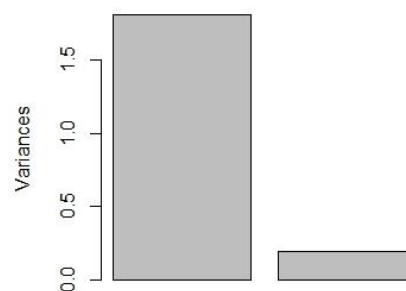
Omdat zowel de *lengte in woorden van het topologische middenveld* als de *lengte in woorden van het cognitieve middenveld* numerieke ratiovariabelen zijn, hebben we een Pearson correlatietest uitgevoerd. Het resultaat is een correlatiecoëfficiënt van 0,874 ($p < 2,2e^{-16}$). Deze correlatiecoëfficiënt geeft aan dat er een zeer sterk positief lineair verband bestaat tussen de lengte van de twee types middenveld. We stellen het lineaire verband ook visueel voor in het volgende spreidingsdiagram:



Figuur 37: Spreidingdiagram *lengte* (in woorden) *topologisch* en *cognitief middenveld*

Een PCA is een tweede manier om een correlatie tussen het topologische middenveld en het cognitieve middenveld te visualiseren. Net zoals bij de vorige PCA's gebruiken we de log2-transformaties van het topologische en cognitieve middenveld.

Het staafdiagram in figuur 38 geeft aan dat de meeste variabiliteit in de dataset door één PC verklaard wordt. In tabel 22 lezen we over hoeveel variabiliteit het precies gaat, namelijk 90,1%. We kunnen de twee oorspronkelijke variabelen dus reduceren tot één hoofdcomponent.



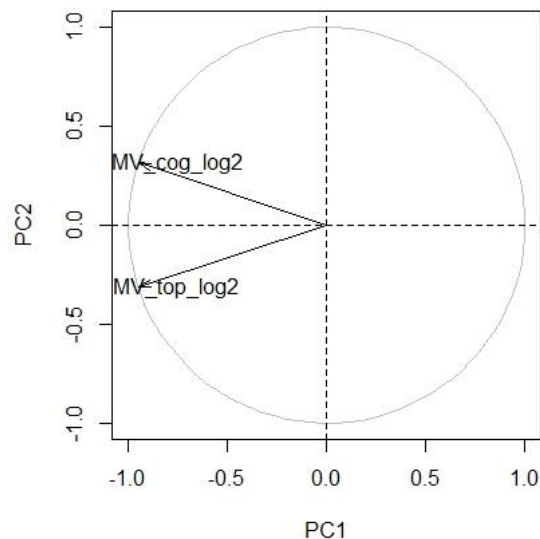
Figuur 38: De relatie tussen *het topologische middenveld* en *het cognitief congruerende middenveld*: de relatieve variantie (y-as) per PC (x-as)

Belang van de componenten		
	PC1	PC2
Standaardafwijking	1.342	0.445
Variantie (proportioneel)	0.901	0.099
Cumulatieve variantie	0.901	1.000

Tabel 22: De relatie tussen *het topologische middenveld* en *het cognitief congruerende middenveld*: overzicht van de standaardafwijking, de variantie en de cumulatieve variantie van de 2 hoofdcomponenten

Figuur 39 toont de correlatiecirkel. Hoewel de pijlen van de twee operationalisaties van het middenveld niet volledige overlappen, wijzen ze wel in dezelfde richting. De hoek tussen de twee variabelen is ook klein genoeg om te besluiten dat de twee

operationalisaties van het middenveld positief gecorreleerd zijn. Het heeft m.a.w. weinig zin om beide operationalisaties aan een multivariate analyse toe te voegen. Net zoals bij de verschillende operationalisaties van syntactische complexiteit selecteren we het best één operationalisatie van het middenveld. We opteren voor het cognitief congruerende middenveld. Het idee dat taalgebruikers de complexiteit van het middenveld willen beperken om sneller het werkwoord te uiten en zo de cognitieve verwerkingskost te beperken, werkt vooral vanuit de veronderstelling dat ze het (getal van het) subject moeten onthouden tot ze het werkwoord bereiken. Het subject en werkwoord zijn namelijk formeel en semantisch met elkaar verbonden, waardoor de taalgebruiker bij het uitspreken of schrijven van het werkwoord zich nog moet herinneren of het subject enkelvoudig of meervoudig was.



Figuur 39: Correlatiecirkel van het *topologische middenveld (log2)* en het *cognitief congruerende middenveld (log2)* op basis van PC1 (x-as) en PC2 (y-as)

4.1.3 Discoursstatus

Operationalisatie1 bekendheid: definietheid

- Hoe geoperationaliseerd?
 - Definiet versus indefiniet (6 observaties NA)
- Type variabele en meetniveau:
 - categorische variabele met nominaal meetniveau

Operationalisatie 2 bekendheid: toegankelijkheid

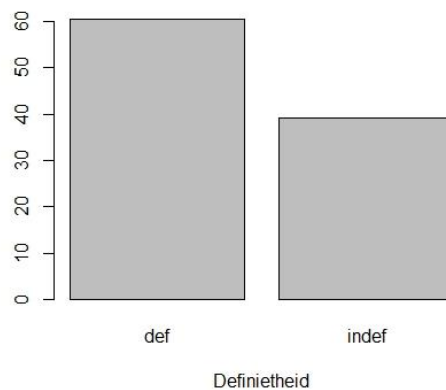
- Hoe geoperationaliseerd?
 - 8 levels van toegankelijke informatie naar nieuwe informatie
- Type variabele en meetniveau:
 - categorische variabele met ordinaal meetniveau

Informationeel belang

- Hoe geoperationaliseerd?
 - (a) aantal keer dat de referent in de voorcontext vermeld wordt = **informationeel belang: voorcontext**
 - (b) aantal keer dat de referent in de nacontext vermeld wordt = **informationeel belang: nacontext**
- Type variabele en meetniveau:
 - discrete numerieke variabelen met ratiomeetniveau

4.1.3.1 De discoursstatus op basis van *definietheid*

De discoursstatus van de VzC hebben we op meerdere manieren geoperationaliseerd. De eerste operationalisatie *definietheid* is een duidelijk voorbeeld van een categorische variabele met een nominaal meetniveau. Deze variabele kent twee levels of categorieën (*definiet* en *indefiniet*), die we elk een waarde kunnen geven (bijvoorbeeld *definiet* = 1 en *indefiniet* = 2). Deze waarden benoemen de categorieën en er bestaat geen rangorde tussen deze categorieën (*indefiniet* is niet meer of minder dan *definiet*). De meest gebruikte plot om deze categorische variabele visueel voor te stellen is een staafdiagram.

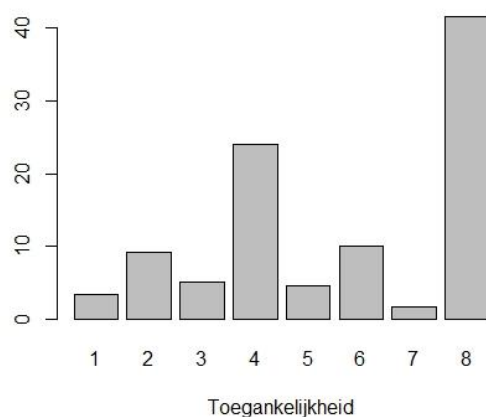


Figuur 40: Staafdiagram van de variabele *definietheid*: overzicht van de relatieve frequentie (y-as) definitie en indefiniete VzC's (x-as)

Figuur 40 geeft een overzicht van het relatieve aantal observaties met een (in)definiete VzC. Op basis van deze figuur stellen we vast dat de dataset iets meer definitie VzC's (ongeveer 60%) dan indefiniete bevat (ongeveer 40%). De som van de staven in figuur 40 is net geen 1 (of 100%). De reden daarvoor is dat we 6 VzC's hebben waarvan we niet kunnen vaststellen of ze definitie of indefiniet zijn. Het gaat meer bepaald om VzC's met een adjectiefconstituent of een adverbiale constituent. Voor de volledigheid voegen we er ook de absolute cijfers aan toe: 1042 zinnen hebben een definitie VzC, 677 zinnen een indefiniete.

4.1.3.2 De discoursstatus op basis van *toegankelijkheid*

De *toegankelijkheid* is de tweede operationalisatie van de bekendheid (en van *discoursstatus* in het algemeen). Zoals in de methodologie duidelijk is geworden, bestaat deze variabele uit 8 verschillende categorieën. Het verschil met *definietheid* is dat de categorieën van *toegankelijkheid* wel een rangorde vertonen. Er is namelijk sprake van een schaal: van afhankelijke en zeer toegankelijke informatie naar nieuwe en ontoegankelijke informatie. We hebben de verschillende categorieën van *toegankelijkheid* ook een numerieke waarde gegeven, ook al kunnen we niet spreken over een ratiovariabele. Er is namelijk geen nulpunt en we kunnen niet beweren dat waarde (of categorie) 3 dubbel zo toegankelijk is als waarde (of categorie) 6. De variabele *toegankelijkheid* heeft daarom een ordinaal meetniveau. Voor een variabele met een ordinaal meetniveau gebruiken we een staafdiagram om de relatieve frequentie visueel voor te stellen.



Figuur 41: Staafdiagram van de variabele *toegankelijkheid* (8 categorieën): overzicht van de relatieve frequentie (y-as) per categorie (x-as)

Figuur 41 toont aan dat de meeste VzC's nieuwe informatie (categorie 8) bevatten. Ook categorie 4 (bestaande uit *onafhankelijke* en *onafhankelijk-collocationele informatie*) is goed vertegenwoordigd. Categorie 7 (*collocationeel-nieuw*) komt duidelijk het minst voor in deze dataset.

Door elke categorie een numerieke waarde te geven, lijkt de variabele *toegankelijkheid* numeriek. Daardoor zouden we ook geneigd kunnen zijn om descriptieve statistische waarden zoals minimum (1), maximum (8), mediaan (6), gemiddelde (5,62) en standaardafwijking (2,32) te berekenen. Toch zeggen deze waarden maar weinig. Het enige wat we wel kunnen afleiden uit deze resultaten is dat het gemiddelde en de mediaan hoger liggen dan bij een normale verdeling (gemiddelde en mediaan zouden dan 4,5 zijn). De hogere resultaten zijn een indicator dat er een rechtse piek is (of een linkse scheefheid; vandaar ook een negatieve waarde van -0,33 voor *skewness*), wat we ook zien in figuur 41 (categorie 8 representeert 41,5% van de dataset).

4.1.3.3 De discoursstatus op basis van *informationeel belang*

De variabele *informationeel belang* meet hoe belangrijk de referent in de context is. De variabele bestaat uit twee verschillende operationalisaties. Enerzijds hebben we rekening gehouden met het aantal keer een referent, i.e. de referent die in de (kern) van de VzC vermeld wordt, in de voorcontext vermeld wordt, anderzijds met het aantal keer een referent in de nacontext voorkomt. Het resultaat zijn twee variabelen: 1. *informationeel belang: voorcontext* en 2. *informationeel belang: nacontext*.

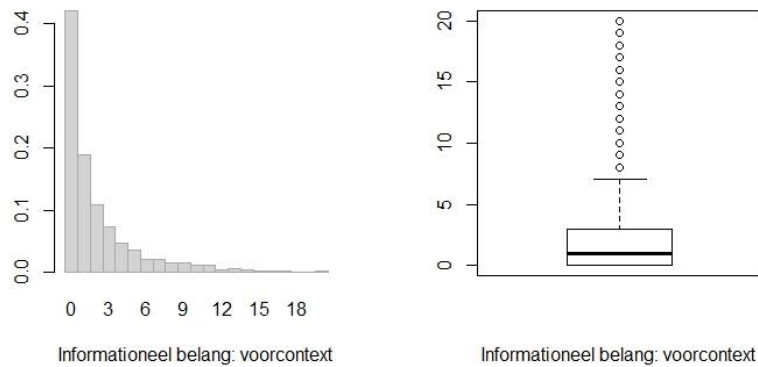
We houden beide operationalisaties gescheiden omdat we te weten willen komen op welke manier de context de positie van VzC's kan bepalen. Er zijn namelijk twee mogelijkheden: ofwel houdt de schrijver rekening met de belangrijkheid van de referent in de voorcontext, ofwel met de belangrijkheid in de nacontext. In het eerste geval zal het aantal vermeldingen in de voorcontext de positie van een VzC beïnvloeden, in het tweede geval het aantal vermeldingen in de nacontext. Anders geformuleerd, we willen aan de hand van deze operationalisaties nagaan of de schrijver voornamelijk bezig is met het belang van de referent op het moment dat die referent geproduceerd wordt, of met het belang van de referent in het daaropvolgende discours.

Beide operationalisaties zijn numerieke variabelen met een ratiomeetniveau (er is een nulpunt, een rangorde en als een referent 4 keer in de context voorkomt, dan is dat dubbel zo vaak als een referent die maar 2 keer in de context vermeld wordt). Daarom gebruiken we voor de frequentie een histogram en voor de vijf-getallensamenvatting een boxplot. In paragraaf 4.1.3.3.1 en 4.1.3.3.2 geven we een descriptief overzicht van deze twee operationalisaties.

4.1.3.3.1 Operationalisatie van *informationeel belang: voorcontext*

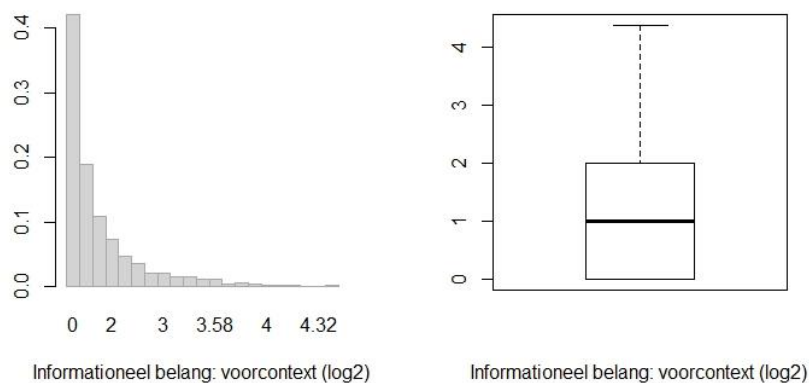
Het histogram (links figuur 42) toont per aantal vermeldingen in de voorcontext (x-as) de relatieve frequentie (y-as). Zo kunnen we uit deze plot afleiden dat de meeste observaties geen eerdere vermelding hebben in de voorcontext (meer dan 40%), één vermelding hebben in de voorcontext (ongeveer 20%) of zelfs twee keer eerder vermeld worden (ongeveer 10%). Deze uitgesproken linkse piek neemt minder snel af naarmate de referent drie of meer dan drie keer in de voorcontext voorkomt. Het histogram wijst dus in de richting van een rechtse scheefheid. Een positieve scheefheidswaarde, namelijk 2,28, bevestigt deze waarneming.

De spreidingsbreedte (van minimum 0 tot maximum 20) lezen we ook af op de y-as in de boxplot (rechts figuur 42). Daarnaast toont de vette lijn in de boxplot de mediaan (= 1,00). Het valt ook op dat er boven de box heel wat *outliers* zijn. Deze *outliers* zullen ervoor zorgen dat het gemiddelde hoger is dan de mediaan (namelijk 2,17) en wijzen tevens op een scheve of asymmetrische verdeling. De standaardafwijking voor deze variabele is 3,27. Deze standaardafwijking geeft aan dat de waarden ver verwijderd zijn van het gemiddelde en dat er dus een grote spreiding is.



Figuur 42: Histogram (links) en boxplot (rechts) van *informationeel belang: voorcontext*

De visuele exploratie van *informationeel belang: voorcontext* illustreert hoe deze variabele niet normaal verdeeld is (Shapiro-Wilk-test: $W = 0,70$, $p < 2,2e^{-16}$) en in extreme mate naar rechts *geskewed* is. We zullen deze variabele daarom logaritmisch transformeren (figuur 43). Net zoals bij de logaritmische transformatie van de *syntactische complexiteit*, kiezen we voor \log_2 (in navolging van Gries (2013: 254)). De spreiding van deze \log_2 -transformatie is kleiner en we stellen geen *outliers* vast in vergelijking met figuur 42. De belangrijkste reden om met deze \log_2 -transformatie verder te werken, is de gereduceerde scheefheid (*skewness* van oorspronkelijk 2,28 naar 0,69).

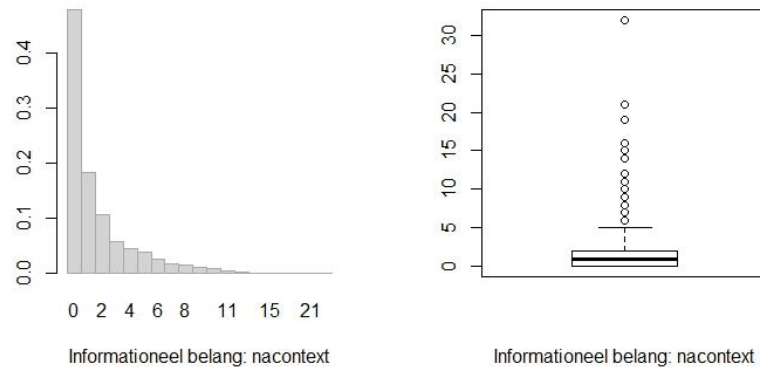


Figuur 43: Histogram (links) en boxplot (rechts) van de \log_2 -transformatie van *informationeel belang: voorcontext*

4.1.3.3.2 Operationalisatie van *informationeel belang: nacontext*

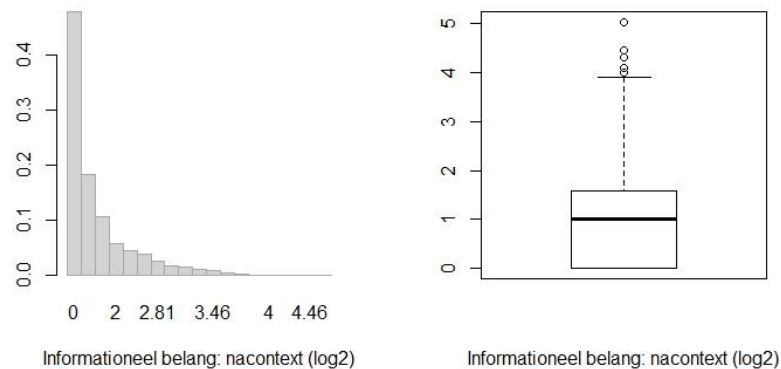
Voor de operationalisatie van *informationeel belang van de nacontext* werken we op dezelfde manier als voor de operationalisatie van de *voorcontext*. Het histogram in figuur 44 illustreert opnieuw een linkse piek (een positieve waarde van 2,96 bevestigt een rechtse scheefheid). Verder geeft het histogram ook aan dat meer dan 45% van de referenten niet meer in de nacontext vermeld wordt, bijna 20% van de referenten nog één keer in de nacontext ter sprake komt en ongeveer 10% nog twee keer. Vanaf drie vermeldingen in de nacontext daalt het histogram minder snel.

De boxplot illustreert eveneens deze asymmetrische verdeling met heel wat outliers boven de box. De mediaan is 1 en het gemiddelde 1,72. Een standaardafwijking van 2,75 wijst ook bij deze numerieke ratiovariabele op een grote spreiding en een asymmetrische verdeling.



Figuur 44: Histogram (links) en boxplot (rechts) van *informationeel belang: nacontext*

Op basis van figuur 44 en de Shapiro-Wilk-test ($W = 0,66$, $p < 2,2e^{-16}$) kunnen we besluiten dat *informationeel belang nacontext*: niet normaal verdeeld is. We zullen deze variabele in de verdere analyse dan ook logaritmisches transformeren (\log_2). Figuur 45 toont aan dat de spreiding van de getransformeerde variabele minder groot is, dat het aantal outliers gereduceerd is en dat bovenal de scheefheid afgenomen is (*skewness* van oorspronkelijk 2,96 naar 0,80).



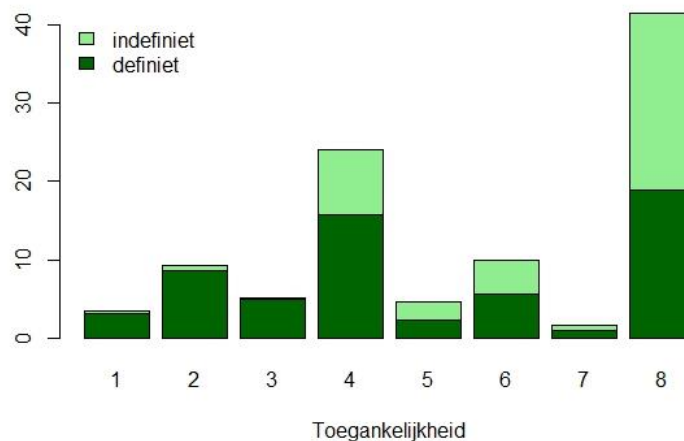
Figuur 45: Histogram (links) en boxplot (rechts) van de \log_2 -transformatie van *informationeel belang: nacontext*

4.1.3.4 De relatie tussen *definietheid* en *toegankelijkheid*

De operationalisaties van *discoursstatus* kunnen we in twee groepen onderverdelen. De eerste twee operationalisaties, *definietheid* en *toegankelijk*, meten de bekendheid van de VzC, de derde operationalisatie, *informationeel belang*, de belangrijkheid van de VzC.

Conceptueel gezien is er dus een verschil op te merken tussen *definietheid* en *toegankelijkheid* aan de ene kant en *informationeel belang* aan de andere kant.

Omdat de operationalisaties *definietheid* en *toegankelijkheid* conceptueel hetzelfde meten, is het goed mogelijk dat ze geassocieerd zijn. Zo zouden we eenvoudigweg kunnen verwachten dat zeer toegankelijke referenten definiëte VzC's zijn en de zeer ontoegankelijke referenten indefiniëte VzC's. Om de relatie tussen deze beide operationalisaties te visualiseren, gebruiken we een gecombineerd staafdiagram (figuur 46).



Figuur 46: Gecombineerd staafdiagram van de variabele *definietheid* (*def/indef*) per level van de variabele *toegankelijkheid*

De variabele *definietheid* bestaat uit twee categorieën (*def(iniet)* en *indef(iniet)*) en de variabele *toegankelijkheid* uit 8 levels. Het gecombineerde staafdiagram (figuur 46) geeft aan dat de eerste drie categorieën van de variabele *toegankelijkheid* (de zogenaamde zeer toegankelijke referenten), zo goed als altijd definiëte constituenten zijn. Vanaf de vierde categorie zien we een verschuiving. In de vierde categorie is de meerderheid van de constituenten nog steeds definiëte (namelijk 65%), maar in de vijfde categorie is de verdeling nagenoeg gelijk (51% definiëte, 49% indefiniëte). In categorieën 6 en 7 zijn er ietwat meer definiëte constituenten dan indefiniëte (namelijk 57% definiëte in categorie 6 en 59% in categorie 7). In categorie 8 tot slot, de nieuwe referenten, is de verdeling omgekeerd en is 54% van de referenten indefiniëte.

Het moge duidelijk zijn dat het staafdiagram onze verwachting voor de zeer toegankelijke referenten min of meer bevestigt. Een zeer toegankelijke referent is dus heel vaak een definiëte constituent. Toch is deze verhouding niet absoluut en bestaan er ook uitzonderingen (bijvoorbeeld in categorie 2 van *toegankelijkheid*, figuur 46).

We veronderstelden ook omgekeerd dat nieuwe referenten voornamelijk indefiniëte VzC's zijn. Het staafdiagram toont echter aan dat een groot aandeel van de nieuwe referenten definiëte is. Wel zien we vanaf de vierde categorie een stijgend aantal indefiniëte constituenten.

Het lijkt er dus op dat relatieve toegankelijkheid goed uitgedrukt kan worden door definiete constituenten, terwijl relatieve ontoegankelijkheid zowel door indefiniete als door definiete constituenten uitgedrukt kan worden.

In een volgende stap willen we statistisch nagaan of de twee variabelen geassocieerd zijn. Omdat de variabelen categoriaal zijn, kunnen we geen correlatietest uitvoeren, zoals een Pearsons correlatietest, noch een meer geavanceerde analyse zoals een PCA.⁸ We kunnen echter wel een Cohens kappa-test uitvoeren die de mate van overeenkomst meet. Op basis van de geschatte kappa kunnen we een uitspraak doen over de eventuele overeenkomst tussen deze twee operationalisaties van discoursstatus. Is de geschatte kappa kleiner dan 0 dan is er geen overeenkomst, tussen 0 en 0,2 een zeer geringe overeenkomst, tussen 0,2 en 0,4 een matige overeenkomst, tussen 0,4 en 0,6 een redelijke overeenkomst, tussen 0,6 en 0,8 een sterke overeenkomst en tussen 0,8 en 1 een bijna perfecte overeenkomst.

De geschatte Cohens kappa voor (de matrix⁹ op basis van) de variabelen *definietheid* en *toegankelijkheid* is 0,14¹⁰ ($Z = 1,07$, $p = 0,14$). Uitgaande van deze kappa-test kunnen we besluiten dat de twee variabelen slechts in een beperkte mate een overeenkomst vertonen. Daarom nemen we de beide operationalisaties mee in ons multivariaat model.

⁸ Pearsons correlatietest veronderstelt een lineaire relatie tussen twee variabelen x en y. Voor een positieve correlatie is de regel 'als x toeneemt dan neemt ook y toe'. In het geval van categoriale variabelen kan je niet zeggen als x toeneemt. De variabele *definietheid* bijvoorbeeld bestaat uit twee levels die niet langer, groter, complexer, enz. worden. Aangezien een PCA zich op de correlatiematrix baseert, kan deze analyse eveneens niet uitgevoerd worden.

⁹ De matrix ziet er als volgt uit:

Levels toegankelijkheid	1	2	3	4	5	6	7	8
definiet	55	148	87	271	41	97	17	326
indefiniet	5	11	3	143	40	75	12	388

¹⁰ Indien we de `Kappa.test()` functie gebruiken in library `fmsb`.

4.1.4 De grammaticale en semantische status van VzC's

Grammaticale status VzC

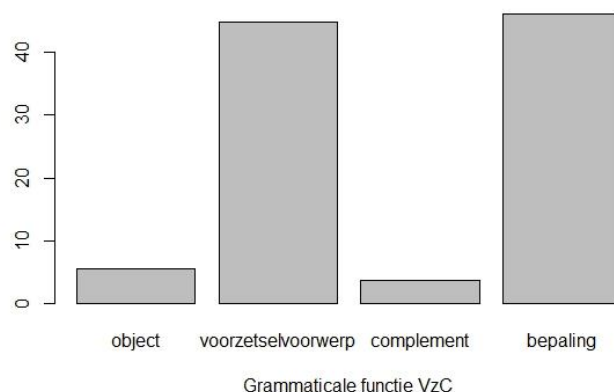
- Hoe geoperationaliseerd?
 - 4 categorieën: object, voorzetselvoorwerp, complement en bepaling
- Type variabele en meetniveau:
 - categorische variabele met nominaal meetniveau

Semantische status VzC

- Hoe geoperationaliseerd?
 - 4 categorieën: categorie 1 (beperking), categorie 2 (locatief), temporeel en fixed
- Type variabele en meetniveau:
 - categorische variabele met nominaal meetniveau

4.1.4.1 De grammaticale status van VzC's

De variabele *grammaticale status van VzC's* bestaat uit 4 categorieën: *objecten*, *voorzetselvoorwerpen*, *complementen* en *bepalingen*. We zouden de verschillende categorieën een waarde kunnen geven (van 1 tot 4) maar deze waarden hebben slechts een oppervlakkige rol. Ze benoemen enkel de verschillende grammaticale functies en geven geen rangorde tussen deze functies weer (voorzetselvoorwerpen zijn niet twee keer objecten, enz.). Net daarom kunnen we niet dezelfde statistische berekeningen uitvoeren als bij ratiovariabelen. In het volgende staafdiagram kunnen we dan ook volstaan met de relatieve observaties per grammaticale categorie. Het valt hierbij op dat objecten en complementen opvallend minder vaak voorkomen (respectievelijk 95 en 63 attestaties of respectievelijk 5,51% en 3,65%) dan voorzetselvoorwerpen en bepalingen (respectievelijk 772 en 795 keer of respectievelijk 44,75% en 46,09%).

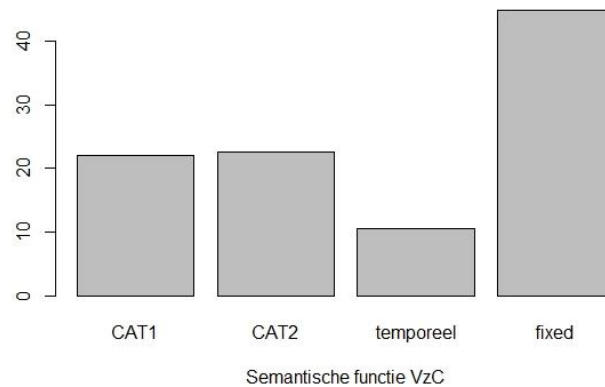


Figuur 47: Staafdiagram *grammaticale status van de VzC*

4.1.4.2 De semantische status van VzC's

Net zoals bij de grammaticale onderverdeling van VzC's bestaat de semantische onderverdeling uit 4 categorieën: *categorie 1* (met *beperking* als de meest frequente semantische klasse), *categorie 2* (met *plaats* als de meest frequente semantische klasse), *temporeel* en *fixed*. We zouden de verschillende categorieën opnieuw een cijfer kunnen geven, maar net zoals bij de grammaticale onderverdeling heeft dit weinig zin. De cijfers hebben geen meerwaarde en benoemen slechts de 4 verschillende categorieën.

We presenteren de variabele semantische functie van de VzC opnieuw in een staafdiagram. De y-as representeert de relatieve frequentie. Het valt hierbij op dat categorieën 1 en 2 min of meer gelijk verdeeld zijn (goed voor elk ongeveer 22%). Ongeveer 10% van de dataset bevat een VzC met een temporele betekenis. De meerderheid (ongeveer 45%) zijn echter voorzetselvoorwerpen met een vast voorzetsel, die we niet kunnen onderverdelen naargelang hun semantische functie.



Figuur 48: Staafdiagram *semantische status van de VzC*

Voor de volledigheid geven we ook hier de absolute frequenties mee. Categorie 1 en 2 komen respectievelijk 379 en 390 keer voor, VzC's met een temporele rol 184 keer en de voorzetselvoorwerpen 772 keer (dat laatste cijfer komt natuurlijk overeen met het aantal voorzetselvoorwerpen in §4.1.4.1).

4.1.5 Het werkwoord

Frequentie van het hoofdwerkwoord

- Hoe geoperationaliseerd?
 - aantal frequenties in het SUBTLEX-NL-corpus
- Type variabele en meetniveau:
 - discrete numerieke variabele met ratiomeetniveau

Semantische classificatie van het werkwoord

- Hoe geoperationaliseerd?
 - 16 levels op basis van de onderverdeling **action**, **process** en **state** (Referentiebestand Nederlands) in combinatie met de **valentie** (intransitief, transitief, ditransitief) en de verplichte aanwezigheid van een **plaats**complement
- Type variabele en meetniveau:
 - categorische variabele met nominaal meetniveau

Type hoofdwerkwoord

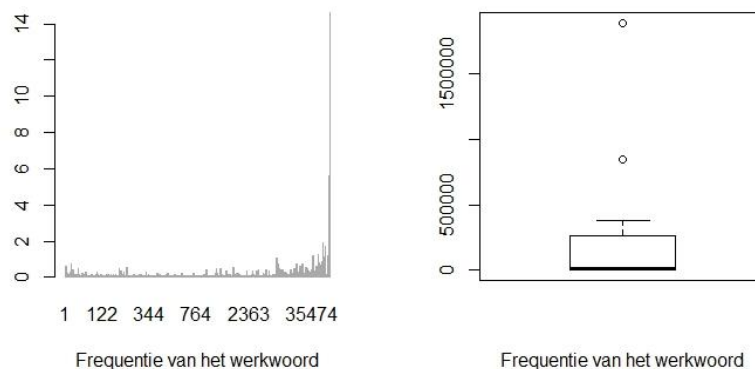
- Hoe geoperationaliseerd?
 - Zelfstandig werkwoord versus koppelwerkwoord
- Type variabele en meetniveau:
 - categorische variabele met nominaal meetniveau

Zinsrelator

- Hoe geoperationaliseerd?
 - 3 levels: 1. enkelvoudige werkwoorden, 2. samengestelde werkwoord en 3. werkwoorden met een partikel, reflexieve werkwoorden of een combinatie van de voorgaande
- Type variabele en meetniveau:
 - categorische variabele met nominaal meetniveau

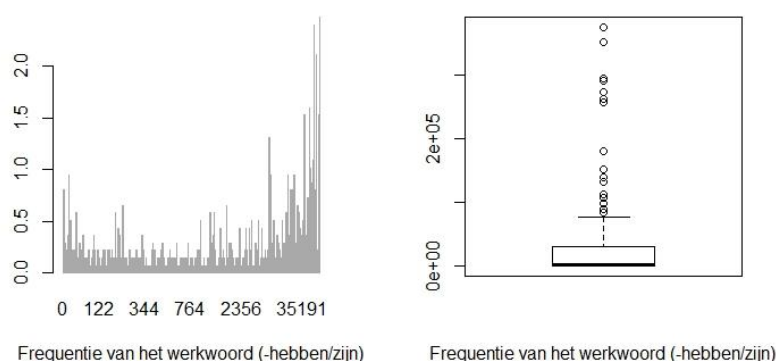
4.1.5.1 De onderverdeling van het werkwoord op basis van de frequentie

Voor de discrete numerieke ratiovariabele *frequentie van het hoofdwerkwoord* gebruiken we een histogram om de frequentie te visualiseren en een boxplot voor de vijf-getallensamenvatting.



Figuur 49: Histogram (links) en boxplot (rechts) van *frequentie van het hoofdwerkwoord*

Het histogram (links figuur 49) geeft een overzicht van de waargenomen (werk)woordfrequenties in het SUBTLEX-NL-corpus (x-as) en het aantal relatieve observaties in onze dataset (y-as). Op basis van het histogram kunnen we besluiten dat er een zeer grote verscheidenheid is aan werkwoorden, die een zeer uiteenlopende frequentiewaarde op basis van het SUBTLEX-NL-corpus hebben. De boxplot geeft aan dat er twee *outliers* zijn: het werkwoord *hebben* (met waarde 845.402 voor frequentie) en het werkwoord *zijn* (met waarde 1.891.414 voor frequentie). Om een idee te krijgen hoe de verdeling eruitziet zonder deze extremen, elimineren we tijdelijk de werkwoorden *hebben* en *zijn* in figuur 50.

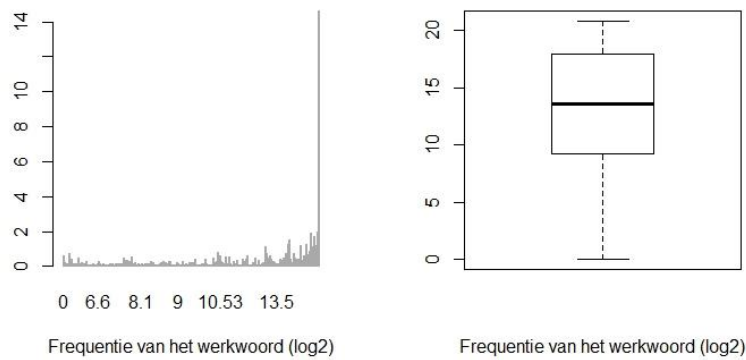


Figuur 50: Histogram (links) en boxplot (rechts) van de *frequentie van het hoofdwerkwoord* (zonder de werkwoorden *hebben*, *zijn*)

Het histogram in figuur 50 illustreert opnieuw de enorme verscheidenheid aan werkwoorden: zowel laagfrequentie werkwoorden als hoogfrequente werkwoorden komen in beperkte mate (tot 2%) aan bod in onze dataset. Volgens de boxplot in figuur 50 zijn de attestaties met een zeer hoge frequentiewaarde outliers.

De spreidingsbreedte van de variabele, met opnieuw de werkwoorden *hebben* en *zijn* toegevoegd, vinden we terug op de x-as van het histogram en op de y-as van de boxplot in figuur 49 (van minimum 0 tot maximum 1.891.414). De vette lijn in de boxplot toont de mediaan. Het lijkt alsof de mediaan samenvalt met 0 maar dat is niet het geval. De mediaan is namelijk 12.671. We vermelden ook nog het gemiddelde 355.834,2 en de standaardafwijking 665.720,3.

Omdat de variabele *frequentie van het werkwoord* niet normaal verdeeld is (Shapiro-Wilk-test: $W = 0,56$, $p < 2,2e^{-16}$) en er een positieve scheefheid van 1,72 is, zullen we deze variabele logaritmisch transformeren. We kiezen daarbij voor \log_2 , in navolging van Gries (2013: 254).

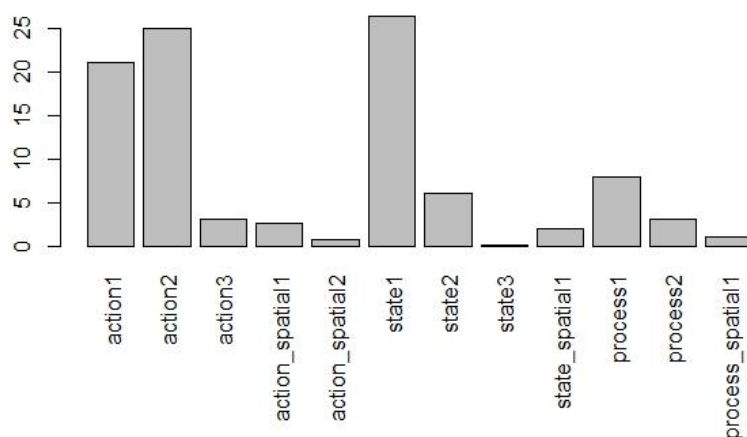


Figuur 51: Histogram (links) en boxplot (rechts) van de log2-transformatie van de variabele *frequentie*

We stellen vast in figuur 51 dat de spreidingsbreedte gereduceerd is van minimumfrequentie 0 tot maximumfrequentie 20,85. Verder toont de boxplot aan dat er geen outliers zijn. De belangrijkste reden om de oorspronkelijke variabele te transformeren is de scheefheid. De scheefheidswaarde van de oorspronkelijke variabele was 1,72. Na de transformatie is die afgenomen tot -0,13.

4.1.5.2 De onderverdeling van het werkwoord op basis van de *semantische classificatie*

De variabele *semantische classificatie van het werkwoord* is een duidelijk voorbeeld van een categorische variabele met een nominaal meetniveau. Deze variabele kent oorspronkelijk 16 levels of categorieën. We stellen deze variabele visueel voor met behulp van een staafdiagram (zie figuur 52).



Figuur 52: Staafdiagram van de variabele *semantische classificatie van het werkwoord*

Figuur 52 geeft een overzicht van het relatieve aantal observaties per semantische categorie van het werkwoord. We stellen vast dat 12 van de 16 oorspronkelijke levels geattesteerd zijn in onze dataset. De afwezige levels zijn: *action_spatial3*, *state_spatial2*,

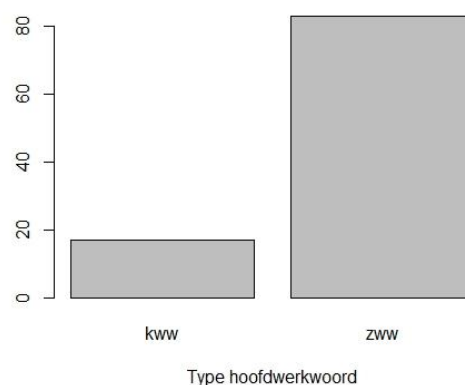
state_spatial3 en *process_spatial2*. We stellen ook vast dat de dataset voornamelijk *action*-werkwoorden bevat (*action1* en *action2*) en *state1*-werkwoorden. Voor de volledigheid voegen we er ook de absolute cijfers toe (in tabel 23)¹¹.

Levels	Absolute frequentie
Action1	364
Action2	431
Action3	53
Action_spatial1	45
Action_spatial2	15
State1	454
State2	106
State3	4
State_spatial1	35
Process1	138
Process2	53
Process_spatial1	19

Tabel 23: Overzicht absolute frequentie per geattesteerde semantische classificatie

4.1.5.3 De onderverdeling van het werkwoord naar het *type hoofdwerkwoord*

De categorische nominale variabele *type hoofdwerkwoord* kan slechts uit twee levels bestaan. Een hoofdwerkwoord is nu eenmaal een koppelwerkwoord of een zelfstandig werkwoord. Net zoals voor de andere categorische variabelen, visualiseren we de relatieve frequentie in een staafdiagram (figuur 53).



Figuur 53: Staafdiagram van de variabele *type hoofdwerkwoord*

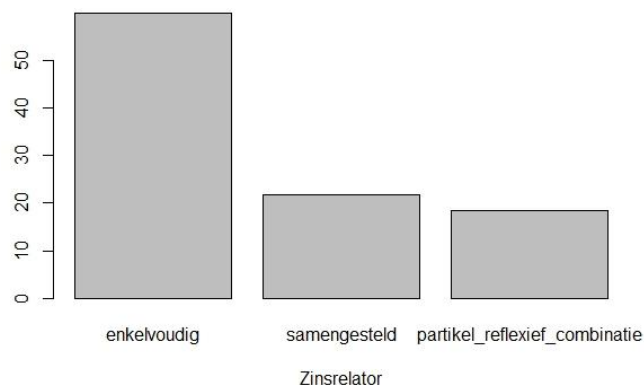
Op basis van figuur 53 stellen we vast dat de dataset opvallend meer zelfstandige werkwoorden bevat dan koppelwerkwoorden: 1431 of 82,96% zinnen hebben een

¹¹ De som van de absolute cijfers is 1717. 8 attestaties kregen het label NA omdat ze niet voorkwamen in het Referentiebestand Nederlands.

zelfstandig werkwoord en 294 of 17,04% een koppelwerkwoord. Dat is op zich niet verwonderlijk, aangezien het aantal koppelwerkwoorden nu eenmaal beperkt is tegenover een brede waaier aan zelfstandige werkwoorden.

4.1.5.4 De onderverdeling van het werkwoord naar de zinsrelator

Voor de laatste operationalisatie van het werkwoord houden we rekening met de *zinsrelator* en meer bepaald met de aanwezigheid van een vast reflexief voornaamwoord, een partikel en een niet-werkwoordelijk deel van een uitdrukking/samengesteld predicaat. Deze onderverdeling verschilt dus van de bovengenoemde, grofmazige onderverdeling in koppelwerkwoorden en zelfstandige werkwoorden. Zo zal het zelfstandige werkwoord *spelen* in de uitdrukking *ergens een rol in of bij spelen* in de categorie *samengesteld werkwoord* opgenomen worden. Indien het werkwoord *spelen* in de letterlijke betekenis wordt gebruikt, zoals in *het kind speelt (met iets)*, dan is het werkwoord wel *enkelvoudig*. De verdeling van de drie categorieën wordt in een staafdiagram (figuur 54) gepresenteerd.



Figuur 54: Staafdiagram van de variabele *zinsrelator*

We stellen vast dat de meeste zinnen een *enkelvoudig werkwoord* hebben (1031 zinnen of 59,77%) en dat het aantal zinnen met een *samengesteld werkwoord* en met een *partikel/reflexief of een combinatie* min of meer gelijk zijn (namelijk 377 zinnen of 21,86% hebben een *samengesteld werkwoord* en 317 of 18,38% een *partikel/reflexief of combinatie*).

4.1.5.5 De relatie tussen het *type hoofdwerkwoord* en de *zinsrelator*

Tot slot vergelijken we de variabele *type hoofdwerkwoord* met de variabele *zinsrelator*. Beide operationalisaties gebeuren op basis van de valentie (i.e. “het geheel aan verbindingsmogelijkheden die het werkwoord heeft op grond van zijn lexicale betekenis” Vandeweghe 2000: 46).

In tabel 24 presenteren we de variabele *type hoofdwerkwoord* (kolom 1) en de variabele *zinsrelator* (rij 1). We stellen vast dat iets meer dan de helft van de koppelwerkwoorden

(ongeveer 58%) in een samengesteld predicaat resulteert (een koppelwerkwoord kan geen deel uitmaken van een werkwoordelijke uitdrukking). In 121 gevallen is het koppelwerkwoord *enkelvoudig* en in 1 geval hebben we een combinatie. Het gaat meer bepaald om de constructie *zich bewust zijn van iets*. Het koppelwerkwoord *zijn* roept in combinatie met het gezegde *bewust* een vast reflexief voornaamwoord *zich* op.

De meerderheid van de zelfstandige werkwoorden resulteren in de categorie *enkelvoudig werkwoord*. In het geval het zelfstandige werkwoord niet enkelvoudig is, komt het vaker voor in een combinatie (316 keer) dan enkel in een uitdrukking of samengesteld predicaat (205 keer).

	enkelvoudig	samengesteld	partikel/reflexief/combinatie	Totaal
koppelwerkwoord	121	172	1	294
zelfstandig werkwoord	910	205	316	1431
Totaal	1031	377	317	1725

Tabel 24: Vergelijking variabelen *type hoofdwerkwoord* en *zinsrelator*

Om de mate van overeenkomst tussen deze twee variabelen te berekenen, gebruiken we Cohens kappa-test. Het resultaat is -0,5 ($Z = -1,22$, $p = 0,89$)¹², wat betekent dat er geen overeenkomst is. Uitgaande van deze statistische test, nemen we de beide variabelen mee in de multivariate analyse.

¹² Indien we de `Kappa.test()` functie gebruiken in library `fmsb`.

4.2 Een multivariate analyse van de verklarende variabelen

Deze studie wil achterhalen op welke manier de onderzochte variabelen de positie van VzC's kunnen bepalen en verklaren. De mogelijks verklarende mechanismen die we in deze studie onderzoeken zijn: de *complexiteit* van het middenveld, achterveld en de VzC, de *discoursstatus* van de VzC (zowel bekendheid als belang), de *grammaticale en semantische status van de VzC* en het *werkwoord*. Vroeger onderzoek beperkte zich tot een monofactoriële analyse, waarbij telkens het effect van één variabele op de woordvolgorde werd getest. Het is echter heel waarschijnlijk dat een taalfenomeen zoals de woordvolgorde niet door één maar daar meerdere factoren wordt bepaald. Een monofactoriële analyse komt bijgevolg niet overeen met de taalrealiteit:

We live in a multifactorial world in which probably no phenomenon is really monofactorial – probably just about everything is correlated with several things at the same time. This is especially true for language, one of the most complex phenomena resulting from human evolution. (Gries 2013: 247)

Om meerdere variabelen tegelijk te kunnen testen, zullen we een multivariate statistische techniek gebruiken, m.n. een logistische-regressieanalyse. Een logistische-regressieanalyse (zie o.a. Rietveld & Van Hout 2005; Baayen 2008; Gries 2013) maakt het mogelijk om het effect van bovengenoemde variabelen (en interactie-effecten) op de woordvolgordevariatie in kaart te brengen. Deze techniek wordt voornamelijk gebruikt als de afhankelijke variabele binair is. De afhankelijke variabele in dit onderzoek is de *positie van de VzC*; het is een categoriale variabele en er zijn twee mogelijke uitkomsten: *middenveldpositie* of *achterveldpositie*. In dit hoofdstuk lichten we deze analyse verder toe (§4.2.1) en testen we de betrouwbaarheid van het model (§4.2.2). Vervolgens visualiseren we de significante effecten (§4.2.3) en berekenen we de verklarende en voorspellende kracht van het model (§4.2.4). Zoals aangehaald aan het begin van dit hoofdstuk zullen we de resultaten in het volgende hoofdstuk interpreteren.

4.2.1 Logistische regressie

De logistische regressie behoort tot de groep van de zogenaamde *Generalized Linear Models* (afgekort GLM; Nelder & Wedderburn 1972) en is waarschijnlijk de meest populaire techniek om een binaire responsevariabele (zoals *middenveldpositie* versus *achterveldpositie*) op basis van verschillende (predictor)variabelen te modelleren. Naast een GLM berekent ook een LRM (of Logistic Regression Model; Harrell 2001) een binaire logistische regressie. De twee technieken hebben elk hun voordeel (voor een overzicht en

vergelijking van deze twee functies in R, zie Speelman 2014) en zullen beide in deze studie uitgevoerd worden.

In een logistische-regressieanalyse wordt aan elke predictorvariabele een regressiecoëfficiënt toegekend die de relatieve impact op de woordvolgorde weergeeft, rekening houdend met en gecontroleerd voor alle andere predictoren. De coëfficiënt drukt de grootte van de kans uit dat de VzC in achterveldpositie of in middenveldpositie zal voorkomen; een regressiecoëfficiënt geeft dus zowel de sterkte als de richting van een effect aan. De relatie tussen de binaire responsvariabele en de predictorvariabelen wordt in de volgende vergelijking uitgedrukt. Links staat de responsvariabele, waarbij y enerzijds gelijk is aan achterveldpositie en anderzijds aan middenveldpositie, rechts de predictorvariabelen x_1, x_2 , enz.

$$\log\left(\frac{p(y = AV)}{p(y = MV)}\right) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots$$

Deze vergelijking modelleert de responsvariabele y in termen van de log van de odds. Dat betekent concreet dat de log van de kans dat y = achterveldpositie gedeeld wordt door de kans dat y = middenveldpositie. De symbolen α en β staan respectievelijk voor de constante (of het intercept¹³) en voor de regressiecoëfficiënten. Aangezien we slechts een steekproef onderzoeken, moeten de regressiecoëfficiënten geschat worden (aan de hand van de *maximum likelihood estimation* methode, zie ook Harrell 2001, Baayen 2008). Als $\beta > 0$ dan stijgt de probabiliteit op succes (in dit geval achterveldpositie) als de waarde van x stijgt, als $\beta < 0$ dan daalt de probabiliteit op succes als de waarde voor x stijgt, en als $\beta = 0$ dan zijn de respons- en predictorvariabele statistisch onafhankelijk van elkaar. De grootte van β bepaalt de mate waarin de probabiliteit op succes stijgt, daalt of gelijk blijft per eenheid dat x stijgt. Het is m.a.w. een graadmeter voor de impact van de predictorvariabele op de woordvolgorde. Op basis van deze regressiecoëfficiënt en de (asymptotische) standaardfout (ASE) kan ook de significantie bepaald worden. Deze significantiewaarde vertelt ons ook welke predictorvariabelen een impact hebben op de woordvolgorde, waardoor we er eveneens een zicht op krijgen welke variabelen relevant zijn voor de woordvolgordebepaling.

De variabelen die we in paragraaf 4.1 uitgebreid hebben besproken, worden aan het regressiemodel toegevoegd. Voor de *syntactische complexiteit* gaat het meer bepaald om de *lengte van het middenveld*, de *lengte van het achterveld* en de *lengte van de VzC*. We hebben aangetoond dat de verschillende operationalisaties in termen van *woorden*, *lettergrepen* en *inbedding* correleren (met behulp van een correlatietest en een PCA), waardoor we slechts een operationalisatie meenemen. Voor de *discoursstatus* werken we met de variabelen

¹³ Simpelweg de waarde van y voor $x = 0$

definietheid, toegankelijkheid, (informationeel) belang_voor en (informationeel) belang_na. Naast de syntactische complexiteit en discoursstatus hebben we ook de grammaticale en semantische status van de VzC, de frequentie van het werkwoord, de semantische classificatie van het werkwoord, het type hoofdwerkwoord en de zinsrelator geoperationaliseerd. Het effect van deze onafhankelijke variabelen op de positie van VzC's (ook wel de afhankelijke variabele) zijn de hoofdeffecten. Voor alle numerieke frequentievariabelen (*lengte van het middenveld, achterveld, VzC, Belang_voor, Belang_na en frequentie van het werkwoord*) werken we met de logaritmische transformaties (\log_2).

Naast de hoofdeffecten voegen we ook enkele interactie-effecten toe. Een interactie-effect veronderstelt dat het effect van variabele x afhangt van de waarde of levels van variabele y ("the effect of x1 on the logit sometimes differs, depending on the value of x2, and conversely that the effect of x2 on the logit sometimes differs, depending on the value of x1" Speelman 2014: 504). Volgens Gries (2013) is de aanwezigheid van interacties essentieel, omdat ze de hoofdeffecten kwalificeren: "The presence of a significant interaction means you cannot take the main effect of the independent variables in the interaction at face value" (Gries 2013: 253). We voegen deze interactie-effecten ook meteen aan het model toe, "because if you do not include interactions in the model equation, they do not get tested and you don't know whether the interaction(s) would in fact help account for the data much better" (Gries 2013: 256).

Wat de inbreng van interactie-effecten betreft, zijn er twee opties: ofwel voegen we alle mogelijke interacties aan het model toe (ook die interacties waarvan we vooraf niets verwachten), ofwel voegen we enkel de interacties toe die we linguïstisch kunnen motiveren en interpreteren. De eerste methode laat ons toe om nieuwe, soms heel verrassende interacties te ontdekken, maar deze onvoorspelbare interacties zijn vaak moeilijk linguïstisch te interpreteren. Daarom beperken we ons tot die interactie-effecten die we zouden kunnen verwachten. De methode die we gebruiken is een *backward selection* gebaseerd op de significantie en het Akaike Information Criterion (ook wel AIC of *criterion-based approach*) (Gries 2013: 260). Dat betekent concreet dat we eerst alle hoofdeffecten en interacties die we kunnen motiveren, aan het maximale model toevoegen. Vervolgens gaan we na welke hoofdeffecten en interacties uit het model verwijderd kunnen worden zonder het model significant slechter te maken. De hoofdeffecten en de interacties die we behouden, vormen het uiteindelijke model (ook wel het *minimal adequate model* (Gries 2013: 260)).

Alle predictorvariabelen, i.e. de hoofdeffecten en interactie-effecten waarvan we verwachten dat ze een effect zullen hebben op dit type van woordvolgordevariatie in het Nederlands, stellen we voor in tabel 25.

Predictorvariabelen			
Onafhankelijke variabelen/ hoofdeffecten		Interactie-effecten	
Syntactische complexiteit	Lengte middenveld (MV) Lengte achterveld (AV) Lengte VzC	Lengte MV: Lengte AV Lengte MV: Lengte VzC Lengte AV: Lengte VzC	Lengte MV: Definietheid Lengte MV: Toegankelijkheid Lengte MV: Belang_voor Lengte MV: Belang_na
Discoursstatus	Definietheid Toegankelijkheid Belang_voor Belang_na	Definietheid: Belang_voor Definietheid: Belang_na Toegankelijkheid: Belang_voor Toegankelijkheid: Belang_na	Lengte AV: Definietheid Lengte AV: Toegankelijkheid Lengte AV: Belang_voor Lengte AV: Belang_na Lengte VzC: Definietheid Lengte VzC: Toegankelijkheid Lengte VzC: Belang_voor Lengte VzC: Belang_na
Status VzC	Grammaticale status Semantische status		
Werkwoord	Frequentie Semantische classificatie Type hoofdwkwoord Zinsrelator		

Tabel 25: Overzicht predictorvariabelen: hoofdeffecten en interactie-effecten

De hypothesen voor de hoofdeffecten hebben we in hoofdstuk 3 besproken. We sommen ze hieronder puntsgewijs op. Voor de invloed van het werkwoord kunnen we enkel vermoeden dat achterveldpositie de voorkeur geniet bij hoogfrequente of semantisch verbleekte werkwoorden. We baseren ons daarvoor op voornamelijk Engelse literatuur (voor een overzicht en meer informatie zie §3.2.4.4).

1. de syntactische complexiteit van het middenveld zal een positief effect hebben op achterveldpositie.
2. de syntactische complexiteit van de VzC zal een positief effect hebben op achterveldpositie.
3. de syntactische complexiteit van het achterveld zal een negatief effect hebben op achterveldpositie.
4. definiete VzC's verkiezen middenveldpositie en indefiniete achterveldpositie.
5. naarmate de VzC ontoegankelijker of nieuwer wordt, wordt achterveldpositie gestimuleerd.
6. informationeel belangrijkere VzC's verkiezen achterveldpositie.
7. complementen en objecten verkiezen middenveldpositie en voorzetselvoorwerpen achterveldpositie.
8. VzC's van tijd en plaats verkiezen middenveldpositie.

We vermoeden dat de complexiteit van het middenveld, het achterveld en de VzC elkaar kunnen beïnvloeden. Ter illustratie: een lang middenveld zal bijvoorbeeld achterveldpositie stimuleren voor een relatief lange VzC, tenzij het achterveld zelf niet leeg is. In dat geval is het mogelijk dat een relatief lange VzC toch in het middenveld staat. Om deze wisselwerking tussen de complexiteit van de verschillende velden en de VzC na te gaan, voegen we alle interacties tussen de complexiteitsoperationalisaties aan het model toe.

Discoursstatus hebben we op twee manieren geoperationaliseerd. We hebben enerzijds gemeten hoe bekend de referent(en) van de VzC is (*definietheid, toegankelijkheid*) en anderzijds hebben we op basis van de voorcontext en nacontext vastgesteld hoe belangrijk de referent(en) is. We zouden ook een interactie-effect kunnen verwachten tussen bekendheid en belang. Ter illustratie: bekende referenten die tegelijkertijd onbelangrijk zijn, zullen zeker middenveldpositie verkiezen, onbekende referenten die belangrijk zijn achterveldpositie. De wisselwerking tussen bekendheid en belang nemen we ook mee in het maximale model.

Tot slot zouden we kunnen verwachten dat er een wisselwerking is tussen de *complexiteit* en *discoursstatus*. Op basis van de literatuur verwachten we dat indefiniëte VzC's, ontoegankelijke VzC's en belangrijke VzC's in het achterveld zullen staan, maar het is evengoed mogelijk dat de complexiteit van het achterveld deze trend afremt. Ook de lengte van het middenveld en de VzC zouden de verwachte effecten voor *discoursstatus* kunnen versterken of afzwakken. Er hangt dus momenteel een waas van onduidelijkheid rond deze twee variabelen (i.e. *complexiteit* en *discoursstatus*) en deze onduidelijkheid is ook zichtbaar in de literatuur. Enerzijds vermeldt de ANS dat de informatiewaarde van de VzC (of de *discoursstatus*) en de *complexiteit* elkaar al dan niet versterken, maar is de manier waarop deze variabelen samenwerken of elkaar tegenwerken niet duidelijk. Braecke (1990), aan de andere kant, vermeldt in zijn empirische onderzoek naar de positie van VzC's dat *syntactische complexiteit* ondergeschikt is aan *discoursstatus* (cf. het Code-Quantity Principle van Givón 1983). Op basis van deze laatste vaststelling zouden we dus kunnen verwachten dat de twee variabelen correleren en dat slechts één variabele significant bijdraagt aan het model, namelijk *discoursstatus*. Om de wisselwerking tussen *complexiteit* en *discoursstatus* te testen, zullen we de interacties van de verschillende *lengtes* met *definietheid, toegankelijkheid, belang_voor* en *belang_na* aan het model toevoegen.

4.2.1.1 Van het maximale model naar het uiteindelijke model

Nu het duidelijk is welke predictorvariabelen we meenemen in het model, fitten we een maximaal logistisch-regressiemodel op onze data. Dat betekent concreet dat we alle hoofdeffecten en alle interactie-effecten uit tabel 25 toevoegen. In onderstaande zoektocht van een maximaal model naar een minimaal, uiteindelijk model vertrekken we

vanuit een GLM, maar we hebben dezelfde redenering en werkwijze ook met een LRM uitgevoerd.

De predictoren die een significant effect hebben, worden in tabel 26 samengevat.¹⁴

Predictoren	β	ASE	Z	Pr(> z)
Lengte_AV_log2	0,22	0,26	0,84	0,40
Lengte_VzC_log2	1,19	0,52	2,28	0,02
Toegankelijkheid (van 1 tot 8)	0,15	0,14	1,07	0,28
Belang_na_log2	-0,58	0,26	-2,25	0,02
PP_semant.class CAT2 (locatief)	-0,69	0,19	-3,57	3,57e-04
PP_semant.class fixed	0,07	0,31	0,24	0,81
PP_semant.class temporeel	-1,97	0,25	-7,81	5,72e-15
ww_type (zww)	-0,95	0,25	-3,87	1,10e-04
ww_zinsrelator partikel_reflexief_combinatie	0,69	0,18	3,75	1,76e-04
ww_zinsrelator samengesteld	0,70	0,20	3,54	4,08e-04
Lengte_AV_log2:Lengte_VzC_log2	-0,20	0,07	-2,75	0,01
Toegankelijkheid:Belang_na_log2	0,07	0,03	2,08	0,04

Tabel 26: Logistische regressie van de toename van achterveldpositie tegenover middenveldpositie voor het journalistieke genre van het DPC corpus (n = 1711, AIC = 1798,3)

Indien een interactie-effect een significant effect heeft op de woordvolgorde, dan behouden we alle hoofdeffecten van de interactie in ons model, want “you cannot delete an insignificant predictor B if the interaction A:B is significant” (Gries 2013: 260). Dat is ook de reden waarom we de hoofdeffecten *lengte van het achterveld* en *toegankelijkheid* in tabel 26 weergeven.

De hoofdeffecten die in het maximale model geen significant effect hebben op de woordvolgorde en die geen deel uitmaken van een interactie-effect zijn: *lengte middenveld*, *definietheid*, *belang_voor*, *grammaticale status van de VzC*, *frequentie van het werkwoord* en *de semantische classificatie van het werkwoord*. De meeste interactie-effecten die we opgenomen hebben, beïnvloeden de woordvolgorde niet. In een volgende stap elimineren we de niet-significante interactie-effecten en houden we de niet-significante hoofdeffecten nog even in het model. De kans bestaat namelijk dat door de eliminatie van een niet-significant interactie-effect een hoofdeffect toch significant wordt. In een regressiemodel beïnvloedt een interactie-effect X:Y de coëfficiënten van de onderliggende hoofdeffecten X en Y (we spreken in de literatuur over conditionele effecten in plaats van hoofdeffecten). Vervolgens testen we ook of dit nieuwe model (dus zonder de niet-significante interacties) significant verschilt van het maximale model. We

¹⁴ 14 observaties werden geëlimineerd als gevolg van ontbrekende waarden. We werken daarom voor de volgende logistische-regressiemodellen met een dataset van 1711 zinnen in plaats van de oorspronkelijk 1725 zinnen. Het gaat meer bepaald om 6 NAs voor de variabele *definietheid* en 8 NAs voor de variabele *semantische classificatie van het werkwoord*.

gebruiken daarvoor ANOVA. Is het verschil significant, dan mogen we de interactie-effecten niet zomaar elimineren. In dat geval is het nieuwe model significant verschillend van het maximale model. Is het verschil echter niet significant, dan is het gereduceerde model niet significant ‘slechter’ dan het maximale model. De p-waarde van de ANOVA-test is 0,54. We kunnen bijgevolg het maximale model vervangen door het gereduceerde model. Deze beslissing wordt ook bevestigd als we de AIC-waarde van de twee modellen vergelijken. De AIC van het maximale model was 1798,3 en is in het nieuwe model gedaald tot 1780,1. Als twee modellen de data (bijna) even goed verklaren, dan is het model met de laagste AIC het best (“if two models explain data equally well, then the model with fewer predictors will have a smaller AIC. Thus, in this approach, a predictor can be added to, or deleted from a model if that lower the AIC” (Gries 2013: 261). De significante predictoren van het gereduceerde presenteren we in tabel 27.

Predictoren	B	ASE	Z	Pr(> z)
Lengte_MV_log2	0,22	0,07	3,21	1,35 ^e -03
Lengte_AV_log2	0,00	0,13	0,02	0,99
Lengte_VzC_log2	1,06	0,10	10,39	< 2^e-16
Toegankelijkheid (van 1 tot 8)	0,11	0,04	2,54	0,01
Belang_voor_log2	0,14	0,08	1,78	0,07
Belang_na_log2	-0,44	0,14	-3,17	1,50e-03
PP_semant.class CAT2 (locatief)	-0,71	0,19	-3,66	2,50e-04
PP_semant.class fixed	0,04	0,30	0,15	0,88
PP_semant.class temporeel	-1,98	0,25	-7,91	2,61e-15
ww_type (zww)	-0,93	0,24	-3,85	1,20e-04
ww_zinsrelator partikel_reflexief_combinatie	0,69	0,18	3,83	1,30e-04
ww_zinsrelator samengesteld	0,69	0,20	3,51	4,44e-04
Lengte_AV_log2:Lengte_VzC_log2	-0,16	0,07	-2,41	0,02
Toegankelijkheid:Belang_na_log2	0,10	0,02	3,87	1,10e-04

Tabel 27: Logistische regressie van de toename van achterveldpositie tegenover middenveldpositie voor het journalistieke genre van het DPC corpus na eliminatie van de niet-significante interactie-effecten (n = 1711, AIC = 1780,1)

Tabel 27 toont enkele grote verschillen in vergelijking met tabel 26. Ten eerste hebben de predictorvariabelen *lengte van het middenveld* en *toegankelijkheid* nu wel een significant effect op de positie van VzC's (wat in het maximale model niet het geval was). Ten tweede heeft de variabele *belang_voor* een borderline effect (p = 0.07).

In een laatste stap elimineren we de niet-significante hoofdeffecten *definietheid*, *grammaticale status van de VzC*, *frequentie van het werkwoord* en *semantische classificatie van het werkwoord*. Deze hoofdeffecten werden niet gepresenteerd in tabel 27 omdat ze geen significant effect hadden op de woordvolgorde en omdat ze geen deel uitmaakten van een significant interactie-effect. Ze waren echter wel nog aanwezig in de logistische-regressieanalyse. Het laatste, minimale model vergelijken we opnieuw met behulp van ANOVA met het vorige model. De p-waarde is 0,4, wat betekent dat het minimale model

niet significant afwijkt van het vorige model. De AIC van het laatste model is ook lager in vergelijking met het vorige model, namelijk 1765,9.

Predictoren	β	ASE	z	Pr(> z)	OR
Lengte_MV_log2	0,25	0,06	3,90	9,46^{e-05}	1,28
Lengte_AV_log2	0,01	0,13	0,07	0,94	
Lengte_VzC_log2	1,06	0,10	10,49	< 2^{e-16}	2,88
Toegankelijkheid (van 1 tot 8)	0,11	0,04	2,70	0,01	1,12
Belang_voor_log2	0,13	0,08	1,72	0,09	
Belang_na_log2	-0,44	0,14	-3,20	1,39^{e-03}	0,64
PP_semant.class CAT2 (locatief)	-0,90	0,17	-5,27	1,40^{e-07}	0,41
PP_semant.class fixed	0,08	0,16	0,50	0,62	
PP_semant.class temporeel	-2,12	0,24	-8,86	< 2^{e-16}	0,12
ww_type (zww)	-0,80	0,18	-4,31	1,66^{e-05}	0,45
ww_zinsrelator partikel_reflexief_combinatie	0,67	0,16	4,19	2,83^{e-05}	1,95
ww_zinsrelator samengesteld	0,76	0,18	4,17	3,00^{e-05}	2,14
Lengte_AV_log2:Lengte_VzC_log2	-0,17	0,07	-2,52	0,01	0,84
Toegankelijkheid:Belang_na_log2	0,09	0,02	3,83	1,28^{e-04}	1,10

Tabel 28: De minimale logistische regressie van de toename van achterveldpositie tegenover middenveldpositie voor het journalistieke genre van het DPC corpus (n = 1711, AIC = 1765,9)

Voor de volledigheid presenteren we ook de R-output van zowel de GLM als de LRM:

```
Call:
glm(formula = PP_positie ~ Lengte_MV_log2 + Lengte_AV_log2 +
  Lengte_VzC_log2 + Lengte_AV_log2:Lengte_VzC_log2 + Toegankelijkheid +
  Belang_voor_log2 + Belang_na_log2 + Toegankelijkheid:Belang_na_log2 +
  PP_semant.class + ww_type + ww_zinsrelator, family = binomial,
  data = data)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.4826  -0.8493   0.3104   0.8180   2.4627

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    -1.928678    0.397833  -4.848 1.25e-06 ***
Lengte_MV_log2    0.247727    0.063453   3.904 9.46e-05 ***
Lengte_AV_log2    0.009842    0.132945   0.074 0.940983
Lengte_VzC_log2    1.058369    0.100868  10.493 < 2e-16 ***
Toegankelijkheid    0.110317    0.040797   2.704 0.006850 **
Belang_voor_log2    0.134748    0.078306   1.721 0.085290 .
Belang_na_log2   -0.436635    0.136570  -3.197 0.001388 **
PP_semant.classCAT2 -0.896263    0.170228  -5.265 1.40e-07 ***
PP_semant.classfixed  0.081901    0.162868   0.503 0.615058
PP_semant.class temporeel -2.116096    0.238967  -8.855 < 2e-16 ***
ww_typezww      -0.795634    0.184762  -4.306 1.66e-05 ***
ww_zinsrelatorpartikel_reflexief_combinatie  0.667406    0.159409   4.187 2.83e-05 ***
ww_zinsrelatorsamengesteld  0.761791    0.182521   4.174 3.00e-05 ***
Lengte_AV_log2:Lengte_VzC_log2 -0.169054    0.067069  -2.521 0.011716 *
Toegankelijkheid:Belang_na_log2  0.093807    0.024486   3.831 0.000128 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 2340.1  on 1710  degrees of freedom
Residual deviance: 1735.9  on 1696  degrees of freedom
AIC: 1765.9

Number of Fisher Scoring iterations: 5
```

Figuur 55: Output R van de uiteindelijke minimale logistische regressie met behulp van GLM

```

Logistic Regression Model

lrm(formula = PP_positie ~ Lengte_MV_log2 + Lengte_AV_log2 +
  Lengte_VzC_log2 + Lengte_AV_log2:Lengte_VzC_log2 + Toegankelijkheid +
  Belang_voor_log2 + Belang_na_log2 + Toegankelijkheid:Belang_na_log2 +
  PP_semant.class + ww_type + ww_zinsrelator, data = data)

```

		Model Likelihood Ratio Test		Discrimination Indexes		Rank Discrim. Indexes	
obs	1711	LR chi2	604.21	R2	0.399	C	0.823
MV	739	d.f.	14	g	1.824	Dxy	0.646
AV	972	Pr(> chi2)	<0.0001	gr	6.196	gamma	0.646
max deriv	2e-07			gp	0.319	tau-a	0.317
				Brier	0.170		

	Coef	S.E.	wald Z	Pr(> Z)
Intercept	-1.9287	0.3978	-4.85	<0.0001
Lengte_MV_log2	0.2477	0.0635	3.90	<0.0001
Lengte_AV_log2	0.0098	0.1329	0.07	0.9410
Lengte_VzC_log2	1.0584	0.1009	10.49	<0.0001
Toegankelijkheid	0.1103	0.0408	2.70	0.0069
Belang_voor_log2	0.1347	0.0783	1.72	0.0853
Belang_na_log2	-0.4366	0.1366	-3.20	0.0014
PP_semant.class=CAT2	-0.8963	0.1702	-5.27	<0.0001
PP_semant.class=fixed	0.0819	0.1629	0.50	0.6151
PP_semant.class=temporeel	-2.1161	0.2390	-8.86	<0.0001
ww_type=zww	-0.7956	0.1848	-4.31	<0.0001
ww_zinsrelator=partikel_reflexief_combinatie	0.6674	0.1594	4.19	<0.0001
ww_zinsrelator=samengesteld	0.7618	0.1825	4.17	<0.0001
Lengte_AV_log2 * Lengte_VzC_log2	-0.1691	0.0671	-2.52	0.0117
Toegankelijkheid * Belang_na_log2	0.0938	0.0245	3.83	0.0001

Figuur 56: Output R van de uiteindelijke minimale logistische regressie met behulp van LRM

Het uiteindelijke model bevat twee significante interactie-effecten (*Lengte AV:LengteVzC* en *Toegankelijkheid:Belang_na*) en vier significante hoofdeffecten die geen deel uitmaken van de interacties (*Lengte van het middenveld*, *semantische status van de VzC*, *het type hoofdwerkwoord* en de *zinsrelator*).^{15&16} De verschillende predictorvariabelen vinden we ook terug in de eerste kolom van tabel 28 (dit is nl. een kopie van de R-output, afgerond tot twee cijfers na de komma). Tabel 28 geeft een zeer gedetailleerd overzicht hoe we de predictoren moeten lezen, welke positie ze stimuleren en hoe groot het effect van deze predictoren is. In wat volgt, bespreken we de belangrijkste gegevens per kolom. De belangrijkste waarden onder de tabel op basis van de GLM-functie (figuur 55) of boven de tabel op basis van de LRM-functie (figuur 56), komen later aan bod (meer bepaald in §4.2.4).

In de eerste kolom valt het op dat de categorische variabelen opgesplitst zijn naar de waarden (of categorieën) die de variabele kan aannemen, met uitzondering van één waarde of categorie. Ter illustratie, de variabele *PP_semant.class* (de *semantische status van*

¹⁵ De predictor *Belang_voor* zouden we ook uit het model kunnen laten en een nieuw model fitten. Hoewel een model zonder *Belang_voor* niet significant verschilt van een model met *Belang_voor* (ANOVA: $p = 0,28$), stijgt de AIC wel. Op basis van Gries (2013: 261) behouden we het model met de laagste AIC.

¹⁶ We merken ook op dat de intercept significant is. Om een significante intercept te begrijpen, verwijzen we naar Spielman (2014: 508): “The intercept test determines whether the intercept differs significantly from zero. The slope test determines whether the slope differs significantly from zero. At a 95% confidence level, we conclude that intercept or slope differ significantly from zero if $p < 0.05$ ”.

VzC's) kan 4 waarden aannemen: CAT1 (met o.a. beperking), CAT 2 (met o.a. plaats), de temporele klasse en de niet-semantiche *fixed-groep*. In tabel 28 wordt CAT 1 niet weergegeven en krijgen alle andere categorieën een eigen regressiecoëfficiënt. De reden daarvoor is dat categorische variabelen anders geëvalueerd worden in een logistische-regressieanalyse dan numerieke variabelen (zoals *lengte MV*). De regressiecoëfficiënt van een van de categorieën van de categoriale variabele wordt gelijkgeschakeld met 0, i.e. de referentiewaarde. Deze referentiewaarde dient als maatstaf om het effect van de andere waarden op de responsvariabele te evalueren. De referentiewaarden van de verschillende categorische variabelen worden niet opgenomen in het regressiemodel (ze dragen geen informatie want $\beta = 0$). In het uiteindelijke, minimale model is CAT 1 de referentiewaarde van de variabele *PP_semant.class*, *koppelwerkwoord* de referentiewaarde van de variabele *ww_type* (type hoofdwerkwoord) en *enkelvoudige werkwoorden* de referentiewaarde van de variabele *ww_zinsrelator*.

De regressiecoëfficiënten in de tweede kolom hebben een positieve of negatieve waarde, waaruit afgeleid kan worden welke predictoren de kans op achterveldpositie doen toenemen (positief) of afnemen (negatief). De (significante) predictoren die achterveldpositie doen toenemen zijn: de *lengte van het middenveld*, de *lengte van de VzC*, de *toegankelijkheid* van de VzC, de *partikel-, reflexieve werkwoorden (of combinaties)*, de *samengestelde werkwoorden* en de interactie van *toegankelijkheid met belang_na*. De predictoren die achterveldpositie doen afnemen zijn: *belang_na*, de *semantiche categorie 2 (met o.a. plaats)*, de *temporele semantiche klasse*, de *zelfstandige werkwoorden* en de interactie van *lengte achterveld met lengte VzC*. Uit de grootte van de regressiecoëfficiënt blijkt ook hoe sterk de impact op de woordvolgorde is (hoe groter de afstand tot 0, hoe sterker de impact). Daaruit blijkt dat 1. de temporele VzC's de grootste impact hebben ($\beta = -2,12$), gevolgd door 2. de *lengte van de VzC* ($\beta = 1,06$), 3. de *locatieve VzC's* ($\beta = -0,90$), 4. het *type werkwoord* ($\beta = -0,80$), 5. de *samengestelde werkwoorden* ($\beta = 0,76$), 6. de *reflexieve, partikelwerkwoorden en combinaties* ($\beta = 0,67$), 7. het *informationeel belang op basis van de nacontext* ($\beta = -0,44$), 8. de *lengte van het middenveld* ($\beta = 0,25$), 9. de *interactie tussen de lengte van de VzC en de lengte van het achterveld* ($\beta = -0,17$), 10. de *toegankelijkheid* ($\beta = 0,11$) en 11. de *interactie tussen toegankelijkheid en informationeel belang* ($\beta = 0,09$). De relatieve sterkte van de predictoren verschilt en we stellen vast dat de semantiche status de grootste impact heeft op dit type van woordvolgordevariatie.

Elke regressiecoëfficiënt wordt gevolgd door een standaardfout (kolom 3) en een z-waarde (kolom 4). Deze z-waarde wordt berekend op basis van de regressiecoëfficiënt en de standaardfout. Op basis van deze z-waarde wordt vervolgens de p-waarde (kolom 5) berekend die de graad van significantie weergeeft.

Tot slot voegen we in tabel 28 een zesde kolom toe, waarin we de odds ratios presenteren. De odds ratios worden op basis van de regressiecoëfficiënten berekend. We gebruiken daarvoor de exponentiële functie (e^β), waarbij e een constante is die bij

benadering overeenkomt met 2,718282. Deze odds ratios maken het mogelijk om te zeggen dat bij elke stijging van 1 unit in de VzC (concreet betekent dat bij elke stijging van 1 woord), de kans op achterveldpositie stijgt met 2,88 ($= e^{1.06}$). Voor de categorische variabele *PP_semant.class* interpreteren we de odds ratios als volgt: de kans op achterveldpositie versus de kans op middenveldpositie stijgt met 0,41 in het geval de VzC naar een plaats (categorie 2) verwijst in vergelijking met VzC's die naar een beperking verwijzen (of naar een andere semantische mogelijkheid die tot categorie 1 behoort).

Met de (positieve of negatieve) regressiecoëfficiënt en de odds ratios in het achterhoofd zullen we in paragraaf 4.2.3 de predictoren visualiseren. Vooraleer we dit doen, is het echter belangrijk om na te gaan of dit model wel betrouwbaar is en of de onafhankelijke variabelen niet correleren (en bijgevolg dezelfde variatie verklaren).

4.2.2 Multicollineariteit

Multicollineariteit duidt op het bestaan van hoge correlaties tussen de verschillende predictoren. Die correlaties kunnen als gevolg hebben dat de regressiecoëfficiënten onbetrouwbaar en onstabiel zijn, waardoor ook het regressiemodel onbetrouwbaar en bijgevolg moeilijk te interpreteren is. Bovendien kan multicollineariteit in een model voor Type I Errors zorgen. Hoewel we niet voor alle variabelen op voorhand kunnen weten of ze correleren, hebben we in een voorbereidende analyse (§4.1) wel al vastgesteld dat de operationalisaties van complexiteit (*lengte in woorden*, *in lettergrepen* en *het aantal inbeddingen*) correleren. Multicollineariteit is dan ook de reden waarom we slechts één operationalisatie aan het model hebben toegevoegd.

Het is ook bekend dat de aanwezigheid van interacties de kans op multicollineariteitsproblemen kan doen toenemen. We berekenen daarom de *variance inflation factor* voor het uiteindelijke model (met alle predictoren gepresenteerd in tabel 28). Indien de *variance inflation factor* (vif) gelijk is aan 1, is er geen collineariteit. Een vif-waarde > 4 is problematisch volgens Speelman (2014: 528). Szmeccsanyi (2006: 215) vermeldt zelfs dat vif-waarden vanaf 2,5 al een probleem kunnen vormen.

Predictoren	VIF
Lengte_MV_log2	1,08
Lengte_AV_log2	8,91
Lengte_VzC_log2	1,29
Toegankelijkheid (van 1 tot 8)	2,60
Belang_voor_log2	2,41
Belang_na_log2	6,30
PP_semant.class CAT2 (locatief)	1,63
PP_semant.class fixed	1,86
PP_semant.class temporeel	1,30
ww_type (zww)	1,20

ww_zinsrelator partikel_reflexief_combinatie	1,11
ww_zinsrelator samengesteld	1,38
Lengte_AV_log2:Lengte_VzC_log2	9,05
Toegankelijkheid:Belang_na_log2	5,33

Tabel 29: *Variance inflation factor* van het uiteindelijke model

Beide interactie-effecten *lengte_AV:lengte_VzC* en *toegankelijkheid:belang_na* hebben een vif-waarde > 4 , net zoals twee van de hoofdeffecten die deel uitmaken van deze interactie-effecten, namelijk *lengte van het achterveld* en *belang_na*. De hoge vif-waarden voor net die hoofdeffecten en interacties zijn eigenlijk niet verwonderlijk, aangezien het interactie-effect gecorreleerd is met het hoofdeffect. In navolging van Speelman (2014) berekenen we daarom de vif-waarden van een model zonder de interactie-effecten.

Another thing that should be tested routinely is whether there are no correlation patterns among the predictors that are so outspoken that this makes the model less reliable (this is the so-called problem of multicollinearity). We can test this with the command `vif()` which we run on the `lrm` model without interactions. Values higher than four should be considered problematic. If they occur, one of the problematic variables should be removed from the model and the model should be rerun and tested again. (Speelman 2014: 528)

De vif-waarden van alle hoofdeffecten zijn kleiner dan 4 en zelfs kleiner dan 2,5. We kunnen besluiten dat er geen problematische variabelen in het uiteindelijke model aanwezig zijn. De onafhankelijke variabelen in het uiteindelijke model correleren niet (of niet in extreme mate) en het model is betrouwbaar om de woordvolgordevariatie (middenveldpositie versus achterveldpositie) te verklaren.

Predictoren	Vif
Lengte_MV_log2	1,08
Lengte_AV_log2	1,05
Lengte_VzC_log2	1,05
Toegankelijkheid (van 1 tot 8)	2,03
Belang_voor_log2	2,39
Belang_na_log2	1,36
PP_semant.class CAT2 (locatief)	1,61
PP_semant.class fixed	1,85
PP_semant.class temporeel	1,30
ww_type (zww)	1,18
ww_zinsrelator partikel_reflexief_combinatie	1,09
ww_zinsrelator samengesteld	1,37

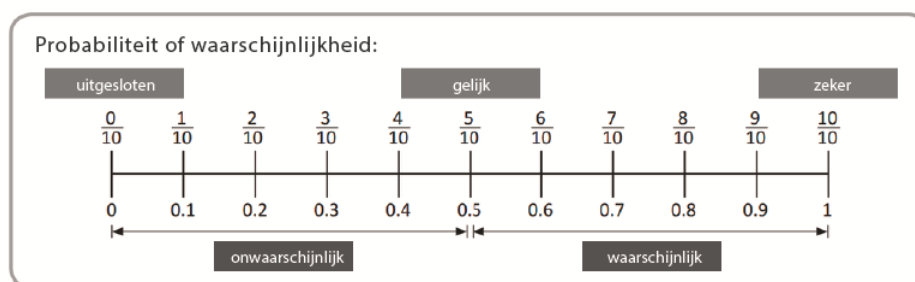
Tabel 30: *Variance inflation factor* van het uiteindelijke model model zonder interactie-effecten

4.2.3 De bespreking en visualisatie van de significante predictoren

In paragraaf 4.2.1 hebben we het uiteindelijke model en de significante predictoren gepresenteerd. In deze paragraaf zullen we nog meer inzoomen op het effect van deze predictoren, door de geschatte probabilmiteit van de predictoren te visualiseren en te bespreken. Waar mogelijk brengen we de bevindingen van onderhavige studie samen met de resultaten op basis van eerder empirisch onderzoek naar de positie van VzC's (§2.2.4)

In de volgende subparagrafen zullen we eerst alle significante hoofdeffecten visualiseren en bespreken en vervolgens de significante interactie-effecten. Het geschatte effect van een predictor dat deel is van een interactie, zoals de *lengte van de VzC*, wordt bepaald en genuanceerd door een andere predictor (in dit geval de *lengte van het achterveld*). Het heeft bijgevolg weinig zin om een hoofdeffect dat deel is van een interactie apart te visualiseren. Voor deze studie betekent dat dat de hoofdeffecten *lengte van de VzC*, *toegankelijkheid* en *belang_na*, die wel een p-waarde kleiner dan 0,05 hebben in tabel 28, niet apart gevisualiseerd en besproken worden.

Voor de plots gebruiken we effect-plots¹⁷. Een effect-plot is interessant omdat het het effect van een predictor in het uiteindelijke logistisch-regressiemodel visualiseert. De x-as presenteert de predictor die we willen visualiseren (of een van de predictoren in het geval van een interactie-effect), de y-as geeft de probabilmiteitsschaal voor de responsvariabele, i.e. achterveldpositie, weer ("the probability scale for effects computed on the logit scale for a binomial GLM" Fox 2003 bijlage pag. 8). De geschatte probabilmiteit is een waarde tussen 0 en 1, waarbij 0 overeenkomt met geen kans op achterveldpositie en 1 met een absolute kans op achterveldpositie. Een waarde van 0,5 geeft aan dat de kans op achterveldpositie en middenveldpositie gelijk is. We stellen deze schaal ook visueel voor in figuur 57.

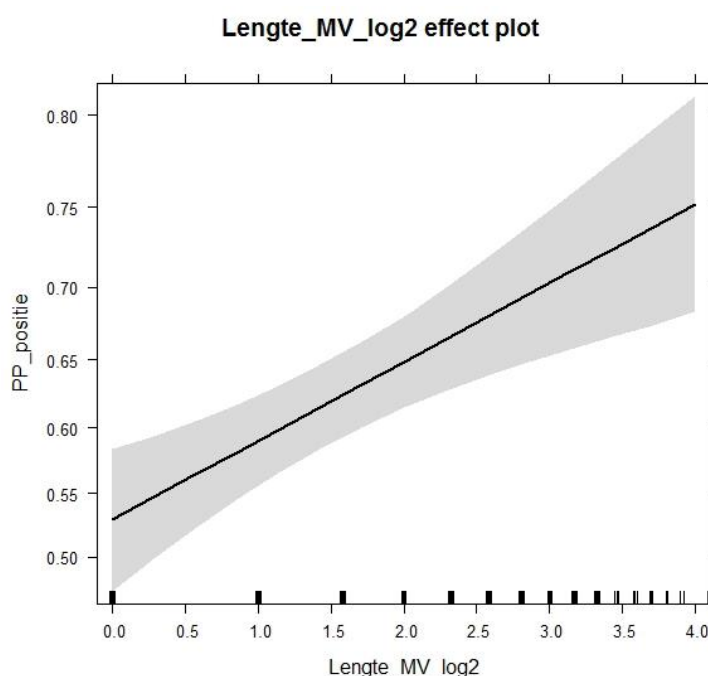


Figuur 57: Probabiliteitsschaal

¹⁷ Effects package in R. Voor meer informatie over dit package, zie Fox & Hong 2009.

4.2.3.1 Hoofdeffect 1: de *lengte van het middenveld*

De effect-plot voor de predictor *lengte van het middenveld* illustreert hoe de lengte van het middenveld de kans op achterveldpositie doet toenemen. De x-as in onderstaande figuur geeft de predictor *Lengte_MV_log2* weer, de y-as de probabilliteit op achterveldpositie.



Figuur 58: Effect plot van *lengte middenveld*

Figuur 58 visualiseert de stijging zoals we die min of meer op basis van de literatuur verwachten (Jansen 1978, 1979; ANS 1997). De kans op achterveldpositie neemt toe, naarmate ook de lengte van het middenveld toeneemt. Het valt echter wel op dat zinnen met een leeg middenveld (0 woorden), een probabilliteit van 0,53 op achterveldpositie hebben. Dat betekent concreet dat de kans op achterveldpositie en middenveldpositie zo goed als gelijk is (in absolute cijfers: 336 zinnen in onze dataset hebben middenveldpositie en 380 zinnen achterveldpositie). Deze vaststelling is op zich wel verrassend. Een leeg middenveld betekent dat er geen overspannen tangconstructie is, wat net middenveldpositie zou kunnen stimuleren. Het valt ook op dat de geschatte probabilliteit voor zinnen met een zeer lang middenveld (bijvoorbeeld zinnen met waarde 3, i.e. 8 woorden tussen de kern van het subject en de persoonsvorm want $\log_2(3) = 8$) ‘slechts’ 0,70 is. In dergelijke zinnen zouden we een absolute voorkeur voor achterveldpositie kunnen verwachten. In het volgende hoofdstuk gaan we hier dieper op in.

Figuur 58 visualiseert ook het 95%-betrouwbaarheidsinterval, wat overeenkomt met de brede grijze waaier rond de geschatte probabilliteit. Concreet betekent dit dat er 95% kans is dat zinnen met een leeg middenveld (van een onbekende dataset) een probabilliteit tussen 0,47 en 0,58 hebben voor achterveldpositie. De kans varieert dus tussen enerzijds geen duidelijk onderscheid tussen middenveldpositie en achterveldpositie (0,47) en

anderzijds een lichte waarschijnlijkheid voor achterveldpositie (0,58). Tabel 31 geeft een overzicht van de verschillende probabiliteitswaarden en het 95%-betrouwbaarheidsinterval (de ondergrens en bovengrens).

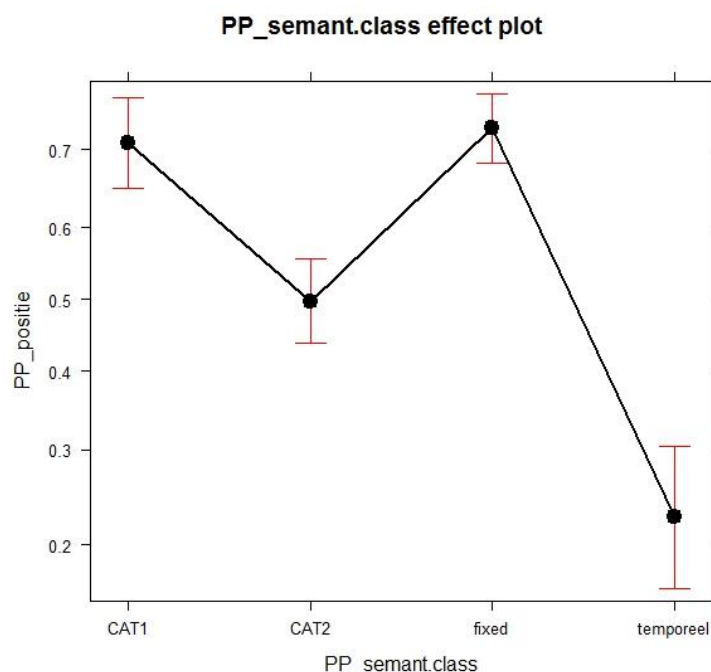
	Lengte van het middenveld log2				
	0	1	2	3	4
Probabiliteit	0,53	0,59	0,65	0,70	0,75
Ondergrens	0,47	0,56	0,61	0,65	0,68
Bovengrens	0,58	0,62	0,68	0,75	0,81

Tabel 31: Overzicht probabiliteit en 95%-betrouwbaarheidsinterval voor de variabele *lengte van het middenveld*

We kunnen samengevat vaststellen dat de lengte van het middenveld in geschreven taal een positieve invloed heeft op achterveldpositie, ook wanneer er tegelijkertijd rekening gehouden wordt met de invloed van andere relevante predictoren. Een schrijver zal een langer middenveld proberen te vermijden door de VzC in het achterveld te plaatsen. De resultaten liggen dus in lijn met het onderzoek van Jansen (1978, 1979) en met wat de ANS beweert.

4.2.3.2 Hoofdeffect 2: de *semantische status van de VzC*

De effect-plot voor de *semantische status van de VzC* leert ons dat categorie2 (met *plaats* als meest voorkomende semantische klasse in deze categorie) en de temporele klasse zich anders verhouden dan categorie 1 (met *beperking* als de meest voorkomende semantische klasse) en de *fixed*-groep. Anders geformuleerd kunnen we op basis van de probabiliteitsschaal vaststellen dat categorie 1 en de *fixed*-groep zeer waarschijnlijk achterveldpositie stimuleren, dat categorie 2 noch middenveldpositie, noch achterveldpositie stimuleert en dat de temporele VzC's voornamelijk middenveldpositie stimuleren. In onderhavig onderzoek staan tijdsbepalingen opvallend vaker in het middenveld in vergelijking met de andere semantische klassen (in absolute cijfers: in deze studie staan 147 tijdsbepalingen in het middenveld, slechts 36 in het achterveld).



Figuur 59: Effect plot van de *semantische classificatie van de VzC*

Figuur 59 toont ook de betrouwbaarheidsintervallen voor de verschillende categorieën. Het 95%-betrouwbaarheidsinterval wordt per categorie in tabel 32 gepresenteerd. De intervallen van CAT1, CAT2 en de temporele klasse overlappen niet. Aangezien CAT1 de referentiewaarde is in het regressiemodel, is er een significant verschil waar te nemen tussen CAT1 en respectievelijk CAT2 en de temporele categorie. Het interval van CAT1 overlapt wel met de *fixed*-categorie, waardoor het effect van de *fixed*-categorie niet-significant is in het logistische-regressiemodel. Het valt dus op dat de enige niet-semantische klasse, namelijk de *fixed*-categorie of ook wel de voorzetselvoorwerpen, niet significant verschilt met de grote groep aan semantische klassen die CAT1 representeert. Hoe we dit moeten interpreteren, blijft koffiedik kijken. Aangezien CAT1 ontzettend veel verschillende semantische klassen bevat, is het goed mogelijk dat er geen duidelijk semantische eenheid aanwezig is in deze categorie, wat ook het geval is bij de *fixed*-groep. VzC's oproepen door een voorzetselvoorwerp zijn namelijk semantisch moeilijk onder te verdelen. Hoe het ook moge zijn, de analyse van de *semantische status van de VzC* toont wel het speciale plaatsingsgedrag van voorzetselvoorwerpen aan, wat onder anderen ook De Schutter (1988) en Hoeksema (2014) al in eerder onderzoek hebben vastgesteld.

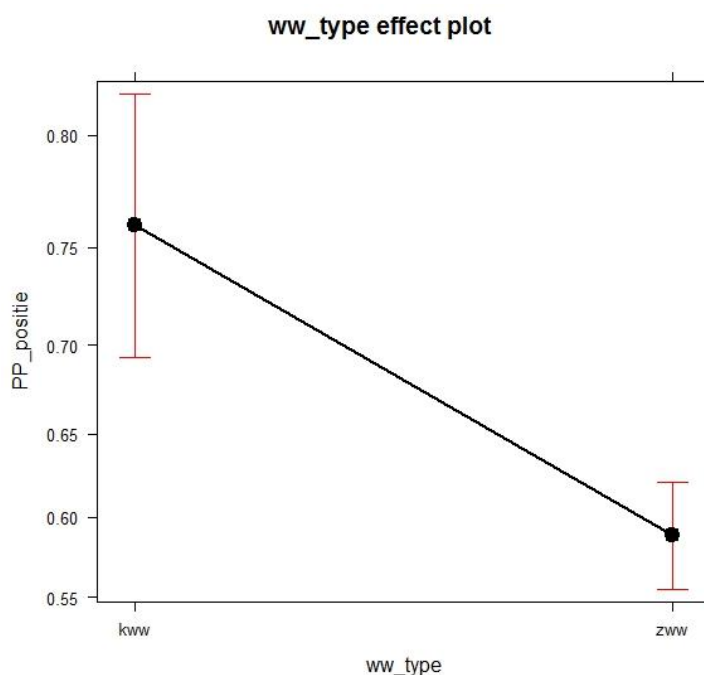
Semantische status van de VzC				
	CAT1	CAT2	Fixed	Temporeel
Probabiliteit	0,71	0,50	0,72	0,23
Ondergrens	0,65	0,44	0,68	0,16
Bovengrens	0,76	0,56	0,76	0,30

Tabel 32: Overzicht probabiliteit en 95%-betrouwbaarheidsinterval voor de variabele *semantische status van de VzC*

De bevindingen voor de semantische klassen op basis van figuur 59 vinden we in zekere zin ook in het onderzoek van Braecke (1990) terug. Braecke stelde namelijk vast dat semantische klassen zoals *beperking*, *doel* en *middel* vaker achteraan staan dan bepalingen van *tijd* en *plaats*. Dit onderzoek kan de voorgaande vaststelling enigszins bevestigen. Beperking en de andere semantische klassen die tot categorie 1 behoren, staan inderdaad opvallend vaker in het achterveld in vergelijking met plaats- en tijdsbepalingen, wat Braecke dus beweert, maar in tegenstelling tot Braecke is er een verschil op te merken tussen de semantisch locatieve klasse (en de andere semantische klassen die tot categorie 2 behoren) aan de ene kant en de semantisch temporele klasse aan de andere kant. Het resultaat is een schaal van zeer weinig kans op achterveldpositie (*tijd*), evenveel kans op achterveldpositie als middenveldpositie (*plaats*) en zeer veel kans op achterveldpositie (*beperking*).

4.2.3.3 Hoofdeffect 3: het *type hoofdwerkwoord*

De effect-plot in figuur 60 geeft aan dat een VzC in combinatie met een koppelwerkwoord meer kans heeft om in achterveldpositie te staan (probabiliteit = 0,76) in vergelijking met een combinatie met een zelfstandig werkwoord (probabiliteit = 0,59). De combinatie van een VzC met een zelfstandig werkwoord heeft wel nog altijd iets meer kans op achterveldpositie dan op middenveldpositie (want probabiliteit = 0,59). De betrouwbaarheidsintervallen overlappen elkaar niet, wat aangeeft dat de twee categorieën significant van elkaar verschillen (wat betreft de geschatte probabiliteit voor achterveldpositie).



Figuur 60: Effect plot voor *type hoofdwerkwoord*

	Type hoofdwerkwoord	
	Koppelwerkwoord	Zelfstandig werkwoord
Probabiliteit	0,76	0,59
Ondergrens	0,69	0,56
Bovengrens	0,82	0,62

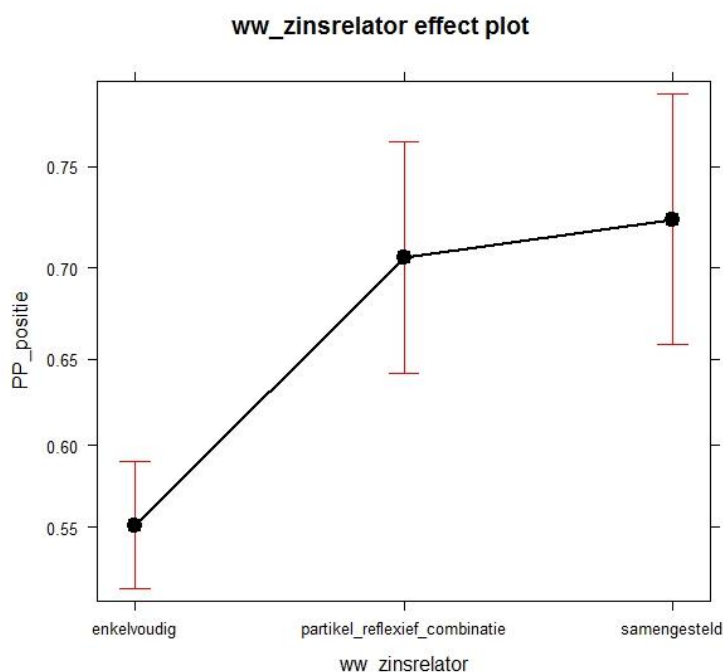
Tabel 33: Overzicht probabiliteit en 95%-betrouwbaarheidsinterval voor de variabele *type hoofdwerkwoord*

Het eerdere empirische onderzoek naar de positie van VzC's heeft het *type werkwoord* (of andere werkwoordsvariabelen) niet meegenomen als een mogelijke verklaring voor deze woordvolgordevariatie. We kunnen daarom bij de bespreking van deze predictor niet terugkoppelen naar paragraaf 2.2.4.

4.2.3.4 Hoofdeffect 4: de zinsrelator

Het laatste significante hoofdeffect is de zinsrelator. We stellen in de effect-plot vast dat de kans op achterveldpositie toeneemt voor samengestelde werkwoorden (probabiliteit = 0.72) en voor de grote groep van partikelwerkwoorden, reflexieve werkwoorden en combinatiegroep, waardoor we kunnen besluiten dat ze significant van elkaar verschillen. De betrouwbaarheidsintervallen van de combinatiegroep en de samengestelde werkwoorden overlappen wel. Deze twee categorieën verschillen m.a.w. niet significant van elkaar in hun geschatte probabiliteit op achterveldpositie¹⁸.

¹⁸ Ze verschillen wel significant van elkaar als we naar de verdeling middenveldpositie en achterveldpositie kijken (in absolute cijfers). De chikwadraattoets bevestigt dat: $\chi^2 = 20.208$, $df = 1$, $p = 6.945e-06$. Dat is ook de reden waarom we deze twee categorieën niet samengenomen hebben (zie methodologie; §3.2.4.3).



Figuur 61: Effect plot van de *zinsrelator*

	Zinsrelator		
	Enkelvoudig	Partikel-/ reflexieve werkwoorden of combinatie	Samengesteld
Probabiliteit	0,55	0,71	0,72
Ondergrens	0,51	0,64	0,66
Bovengrens	0,59	0,76	0,78

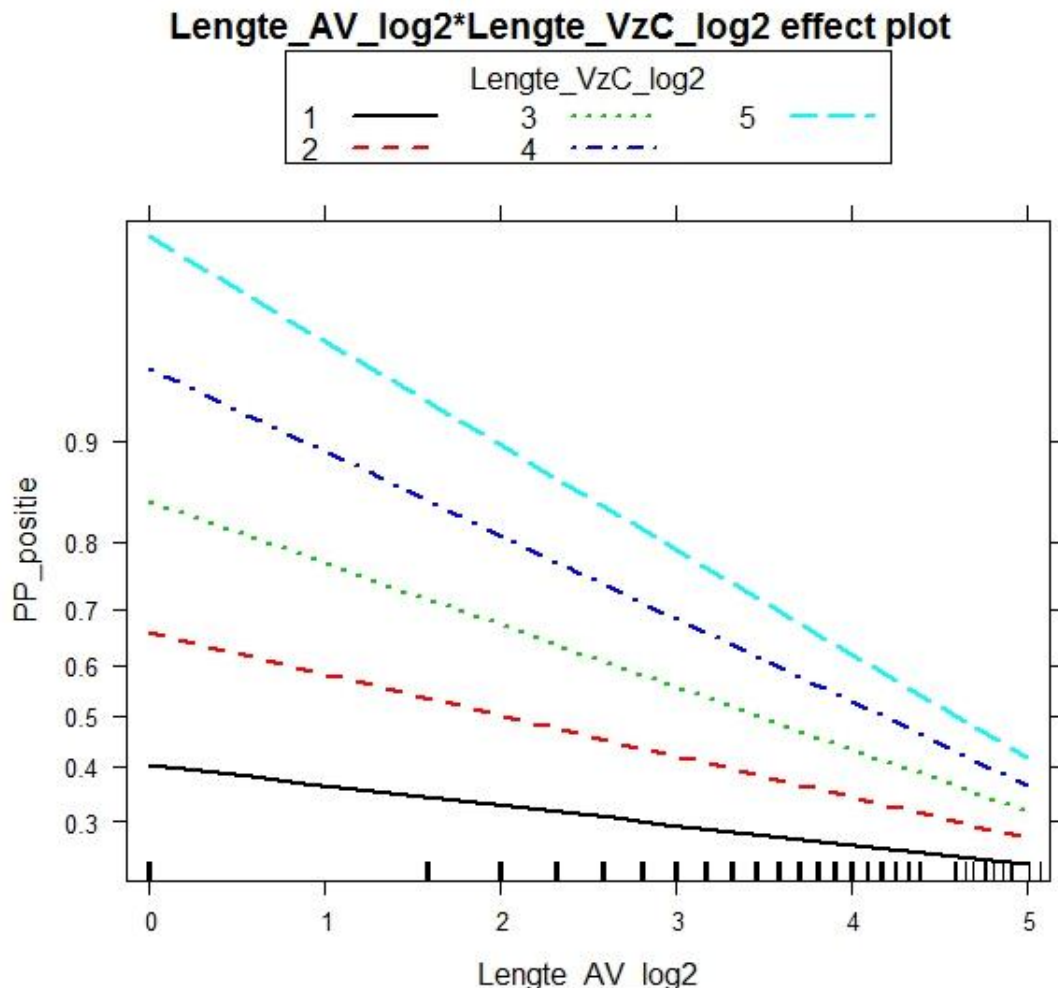
Tabel 34: Overzicht probabiliteit en 95%-betrouwbaarheidsinterval voor de variabele *zinsrelator*

Net zoals bij het type hoofdwerkwoorden kunnen we de resultaten in figuur 61 niet relateren aan eerdere literatuur, aangezien het werkwoord nog niet eerder onderzocht werd voor dit type van woordvolgordevariatie.

4.2.3.5 Interactie-effect 1: de *lengte van het achterveld* en de *lengte van de VzC*

Het eerste significante interactie-effect is een combinatie van de *lengte van het achterveld* met de *lengte van de VzC*. Met behulp van een effect-plot kunnen we deze interactie visualiseren. De x-as in figuur 62 geeft de *lengte van het achterveld* (logaritmisch getransformeerd) weer, de y-as de geschatte probabiliteit voor achterveldpositie. De *lengte van de VzC* (logaritmisch getransformeerd), de andere predictor in het interactie-effect, wordt vervolgens geplotted. De effect-plot maakt daarbij een eigen onderverdeling

voor de lengte van de VzC: lijn 1 = alle zinnen met waarde $[1, 2)^{19}$, lijn 2 = alle zinnen met waarden $[2, 3)$, lijn 3 = alle zinnen met waarden $[3, 4)$, lijn 4 = allen zinnen met waarden $[4, 5)$, lijn 5 = alle zinnen met waarden $[5, 5.615]$. We merken hierbij op dat een waarde 1 niet overeenkomt met 1 woord in de VzC (wat overigens niet kan, aangezien een VzC altijd uit minstens 2 woorden zal bestaan). De variabelen zijn logaritmisch getransformeerd (\log_2), dus waarde 1 komt overeen met 2 woorden ($\log_2(2) = 1$).



Figuur 62: Effect plot van de interactie *lengte van het AV* (\log_2) en *lengte van de VzC* (\log_2)

We stellen in figuur 62 het volgende vast. Ten eerste blijft de volgorde van de lijnen gelijk. Lijn 5 ligt telkens boven lijn 4, lijn 4 boven lijn 3, enz. De lijnen kruisen elkaar niet. Dat betekent concreet dat een langere VzC altijd meer kans heeft om in het achterveld te staan in vergelijking met kortere VzC's. Ten tweede stellen we vast dat de lijnen dalen, naarmate de lengte van het achterveld toeneemt. De daling van deze lijnen gebeurt

¹⁹ De ronde haken geven aan dat de waarde tot het interval behoort. Ter illustratie: voor het eerste interval $[1, 2)$ maakt de waarde 1 en 1,58 deel uit van het interval, waarde 2 niet meer. Het volgende interval $[2, 3)$ is inclusief waarde 2, maar exclusief waarde 3. Tot dit laatste interval behoren ook de waarden 2,32; 2,58 en 2,81.

daarbij niet evenwijdig, maar neemt meer af voor de langere VzC's dan voor de kortere (het lijkt alsof de lijnen in één punt samenkomen). Het achterveld zorgt er dus voor dat de bijna absolute kans op achterveldpositie voor lange VzC's zo goed als verdwijnt indien het achterveld zelf complex is (bijvoorbeeld probabilliteit van 0,98 voor de langste VzC's daalt tot 0,42 als het achterveld zelf lang is).

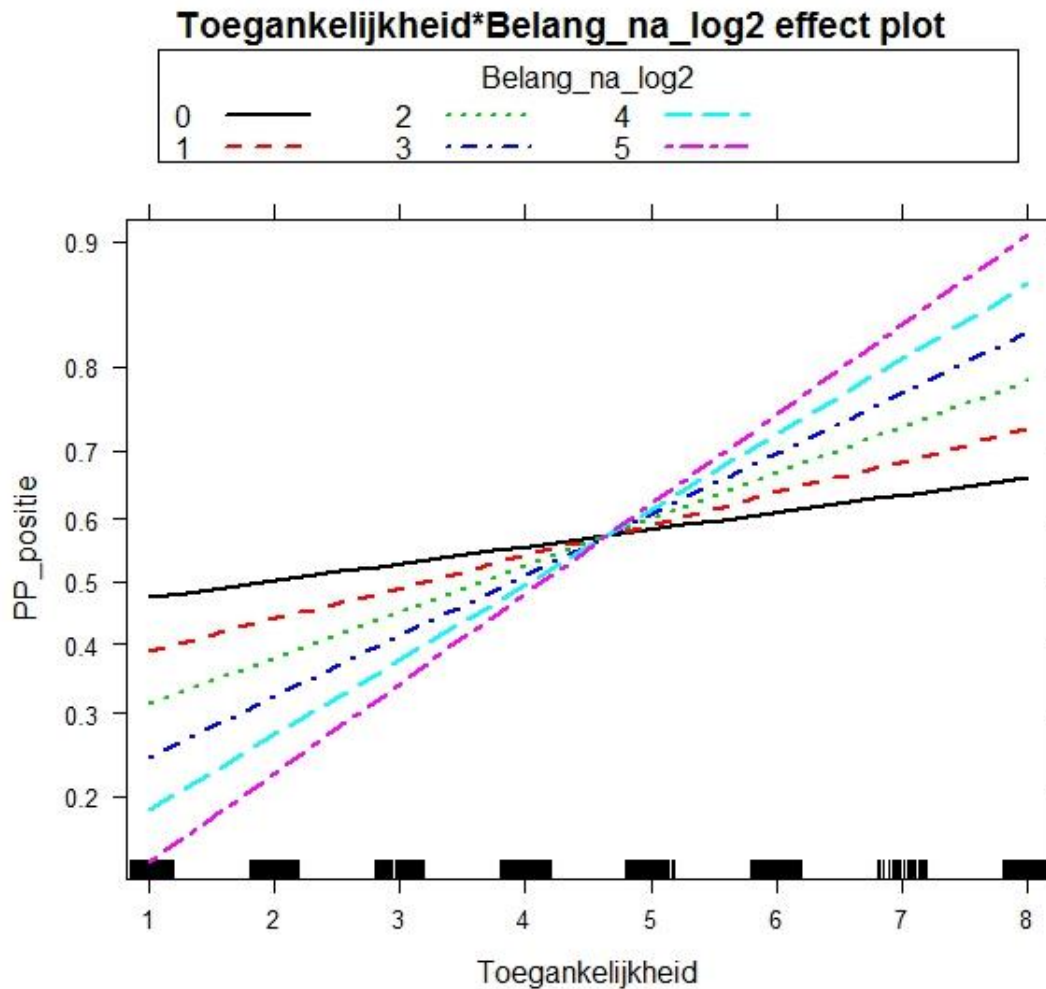
		Lengte VzC_log2				
Lengte AV_log2		1	2	3	4	5
	0	0,40	0,66	0,85	0,94	0,98
	1	0,36	0,58,	0,77	0,89	0,95
	2	0,33	0,50	0,67	0,81	0,90
	3	0,29	0,42	0,56	0,69	0,79
	4	0,26	0,34	0,43	0,53	0,62
	5	0,23	0,27	0,32	0,36	0,42

Tabel 35: Overzicht probabilliteit en 95%-betrouwbaarheidsinterval voor de interactie *Lengte_AV* en *Lengte_VzC*

Op basis van figuur 62 kunnen we concluderen dat de lengte van het achterveld het effect van de lengte van de VzC afremt.

4.2.3.6 Interactie-effect 2: *toegankelijkheid* en *informationeel belang_na*

Het tweede interactie-effect bestaat uit de variabelen *toegankelijkheid* en *belang_na(context)*. Beide variabelen operationaliseren de *discoursstatus*, waarbij de eerste variabele de *bekendheid* van de referent vaststelt, de tweede het *belang* van de referent op basis van het aantal vermeldingen in de context na de geattesteerde zin. De y-as in onderstaande effect-plot geeft de geschatte probabilliteit op achterveldpositie weer, de x-as representeert de 8 levels van de variabele *toegankelijkheid*. Deze levels stellen een rangorde voor van zeer toegankelijke referenten tot zeer ontoegankelijk (en bijgevolg nieuwe) referenten. De laatste predictor *belang_na* wordt opnieuw onderverdeeld door het programma zelf. Dat betekent concreet dat het aantal vermeldingen in intervallen wordt gesplitst en elk interval wordt met behulp van een andere kleur en lijntype geplot. Net zoals bij de verschillende lengtematen, werd ook de frequentievariabele *belang_na* logaritmisch getransformeerd.



Figuur 63: Effect plot van de interactie *toegankelijkheid* en *belang nacontext (log2)*

Op basis van figuur 63 stellen we vast hoe het effect van de variabele *belang_na* beïnvloed wordt door de toegankelijkheid van de referent (of omgekeerd dat de variabele *toegankelijkheid* beïnvloed wordt door het aantal vermeldingen in de nacontext). Voornamelijk VzC's met veel vermeldingen in de nacontext, de zogenaamde belangrijke VzC's (roze lijn in figuur 63), hebben een heel sterk positief lineair effect op achterveldpositie naargelang de graad van toegankelijkheid toeneemt. Tabel 36 toont hoe de geschatte probabiliteit voor deze belangrijke VzC's (want ze zijn nog heel actief in de nacontext) schommelt van geen voorkeur voor achterveldpositie indien de VzC toegankelijk is (0,14) tot een bijna absolute voorkeur voor achterveldpositie indien de VzC nieuw is (0,90). Deze lineaire stijging is veel minder sterk naarmate het belang van de referent daalt. Bij de onbelangrijke VzC's, i.e. in deze studie VzC's met geen of weinig vermeldingen in de nacontext, is er slechts een beperkte lineaire stijging waar te nemen. In het laatste geval schommelt de probabiliteit maar van 0,47 (geen duidelijke voorkeur voor midden- of achterveldpositie) tot 0,60 indien de referent nieuw is (de zwarte lijn in figuur 63).

		Belang_na log2					
Toegankelijkheid		0	1	2	3	4	5
	1	0,47	0,39	0,31	0,24	0,19	0,14
	2	0,50	0,44	0,38	0,32	0,27	0,22
	3	0,53	0,49	0,45	0,41	0,38	0,34
	4	0,56	0,54	0,53	0,51	0,50	0,48
	5	0,58	0,59	0,60	0,61	0,61	0,62
	6	0,61	0,64	0,67	0,70	0,72	0,75
	7	0,64	0,69	0,73	0,77	0,81	0,84
	8	0,66	0,73	0,79	0,83	0,87	0,90

Tabel 36: Overzicht probabilliteit en 95%-betrouwbaarheidsinterval voor de interactie *Toegankelijkheid* en *Belang_na*

Tot slot staan we nog even stil bij de variabele *belang_na* (log2). Op basis van de literatuur verwachten we een lineaire stijging voor belang en achterveldpositie (hoe belangrijker de referent, hoe meer achterveldpositie; *onbelangrijk-voor-belangrijk-regel* zie ook Spielman & Grondelaers 2007), wat we ook zien in figuur 63. Toch is er iets vreemds aan de hand. De regressiecoëfficiënt in tabel 28 was namelijk negatief. Op basis van deze regressiecoëfficiënt zouden we dus verwachten dat de kans op achterveldpositie net afneemt, naarmate het aantal vermeldingen in de nacontext toeneemt en dus naarmate de referent (van de VzC) belangrijker is. Een mogelijke verklaring voor dit tegenstrijdige resultaat is multicollineariteit. In paragraaf 4.2.2 hadden we namelijk al vastgesteld dat de predictor *belang_na* en het interactie-effect *toegankelijkheid:belang_na* een verhoogde waarde voor multicollineariteit hadden. Het probleem met collineariteit is dat er geen unieke oplossing is voor de schatting van een coëfficiënt, waardoor meerdere schattingen mogelijk zijn en het teken (positief/negatief) kan veranderen. De correcte hypothese voor *belang_na* zit bijgevolg in de interactie en het negatieve teken van het hoofdeffect kan wijzen op een ‘restje’ verklaarde variatie. Een tweede verklaring heeft te maken met de effect-plot zelf. De effect-plot geeft namelijk de voorspelde probabilliteit weer en werkt niet op basis van de regressiecoëfficiënt. Wat de verklaring voor de negatieve regressiecoëfficiënt ook moge zijn, we weten dat het hoofdeffect *belang_na* bepaald en beïnvloed wordt door de *toegankelijkheid* en dat dit interactie-effect achterveldpositie motiveert naarmate de referent toegankelijker én belangrijker is.

4.2.4 Verklarende en voorspellende kracht van minimale model

In deze laatste paragraaf gaan we na hoe succesvol het uiteindelijke regressiemodel is. Daarvoor zullen we testen hoeveel van de geobserveerde variatie het regressiemodel kan verklaren en voorspellen. In deze paragraaf zullen we ook naar figuren 55 en 56 verwijzen, of respectievelijk de GLM-output en LRM-output in R, en enkele waarden toelichten die we momenteel nog links hebben laten liggen.

De verklarende waarde van het model kan uitgedrukt worden in de totale reductie in residuele deviantie ten opzichte van een *intercept-only*-model. De residuele deviantie is een maat voor verklaarde variantie. De regel is: hoe minder residuele variantie er overblijft na het fitten van een model, hoe groter de verklarende kracht. De residuele deviantie in het *intercept-only*-model bedraagt 2340,1 (zie ook figuur 55), terwijl de residuele deviantie in het uiteindelijk model gedaald is tot 1735,9.²⁰

Een andere optie om de verklarende kracht van het model vast te stellen, is de ratiotest LR chi² in de LRM-output, ook wel de *model likelihood ratio chi-square* (Speelman 2014: 511). Deze LR chi²-test gaat na of de deviantie in het uiteindelijke model significant kleiner is dan de deviantie in een *intercept-only*-model. De p-waarde van deze test is < 0,0001, wat aangeeft dat het verschil tussen het uiteindelijke model en het *intercept-only*-model statistisch significant is.

Voor de voorspellende kracht van dit model gebruiken we de concordantiemaat of c-maat (Harrell 2001). Deze c-maat vinden we ook terug in figuur 56 of in de LRM-output. De c-maat is een index voor de correlatie tussen de voorspelde probabiliteiten en de geobserveerde waarden van de responsvariabele. De c-maat kan een waarde aannemen tussen 0,5 (geen voorspellende kracht) en 1 (perfect voorspellende kracht). Een waarde hoger dan of gelijk aan 0,80 (of 80%) geeft aan dat het model met alle aanwezige predictoren voldoende voorspellende kracht heeft. Bij een waarde lager dan 0,80 is het aan de onderzoeker om op zoek te gaan naar nieuwe verklaringen. De c-maat voor het uiteindelijk model is 0,823.²¹

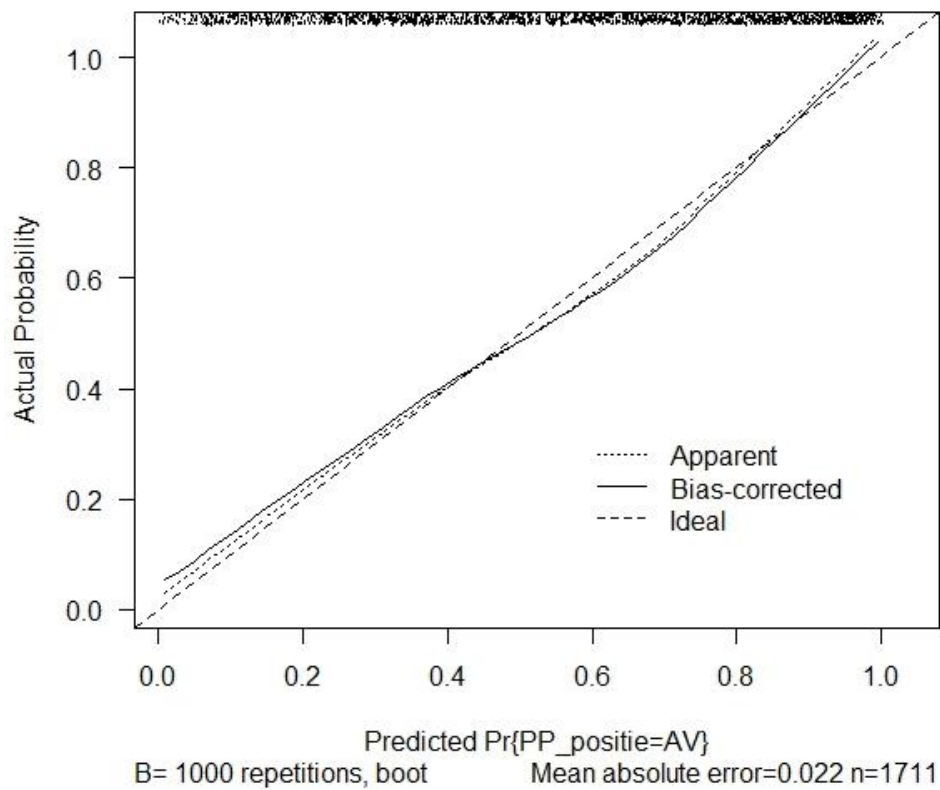
We houden ook rekening met *overoptimisme* of *overfitting* van de c-maat (Harrell 2001: 87-90, 94-97). Het is namelijk mogelijk dat de c-maat hoger is omdat we natuurlijk op eenzelfde dataset testen, dan mochten we op een onafhankelijke dataset testen. Daarom gebruiken we een bootstrapvalidatie (met 1000 bootstrapherhalingen). De c-index is na deze bootstrapvalidatie 0,8175.²² Op basis van deze berekening stellen we vast dat er slechts in een zeer beperkte mate sprake is van *overfitting* en dat deze *overfitting* de verklarende kracht van het uiteindelijke model nauwelijks aantast. Deze beperkte *overfitting* kunnen we ook visueel voorstellen met behulp van een calibratiecurve (opnieuw met 1000 bootstrapherhalingen). Volgens deze calibratiecurve onderwaardeert het uiteindelijke model de lage probabiliteiten (tot ongeveer 0,45), overwaardeert de middelste probabiliteiten (tot ongeveer 0,85) en onderwaardeert opnieuw de hoogste probabiliteiten. Deze onder- en overwaardering gebeurt slechts in beperkte mate en de

²⁰ Daarnaast toont de GLM-output ook de AIC. “The AIC: value is a corrected version of the deviance that has the additional benefit that it can be compared across models (for the same data set and response variable) with different numbers of predictors” (Speelman 2014: 510).

²¹ De c-maat wordt berekend met behulp van de Dxy-maat. Een Dxy van 0,60 komt overeen met een c-maat van 0,80. De Dxy van het uiteindelijk model is 0,646.

²² De gecorrigeerde Dxy = 0,635 dus $c = \frac{1+0,635}{2} = 0,8175$.

gecorrigeerde c-maat met waarde 0,8175 geeft aan dat de verklarende kracht van het uiteindelijke model nog steeds voldoende is.



Figuur 64: Calibratiecurve met 1000 bootstrapherhalingen

4.3 Besluit

In dit hoofdstuk hebben we de verschillende variabelen eerst descriptief geëxploreerd om te weten te komen hoe de variabelen in onze dataset verdeeld zijn. Op basis van deze exploratie hebben we vastgesteld dat de ratiovariabelen niet-normaal verdeeld waren. In navolging van de literatuur hebben we deze variabelen daarom logaritmisch getransformeerd. Daarnaast toonde de exploratie ook aan dat de verschillende operationalisaties van *complexiteit* correleren. Omdat correlerende variabelen voor multicollineariteit kunnen zorgen in een multivariate analyse (waardoor de analyse minder betrouwbaar is), namen we slechts één operationalisatie mee in het verdere onderzoek.

In een tweede stap hebben we een multivariate logistische-regressieanalyse uitgevoerd. Deze analyse toonde aan dat verschillende predictoren een effect hebben op de variabele positie van VzC's. De woordvolgorde wordt dus niet door één maar door meerdere variabelen tegelijkertijd bepaald en verklaard. Dat laatste is nieuw, aangezien eerder empirisch onderzoek de variabelen met behulp van een univariate analyse onderzocht (dus de analyse van telkens één variabele op de woordvolgorde). Door deze multivariate statistische analyse (in combinatie met de methodologische gestrengheid waarmee we te werk zijn gegaan) hebben we bijgevolg een grote stap voorwaarts gezet naar het hoe en waarom van deze woordvolgordevariatie.

Op basis van de analyse kunnen we bevestigen dat de *complexiteit*, de *discoursstatus*, de *semantische status van de VzC* en het *werkwoord* een rol spelen. De analyse onthulde ook twee interessante interactie-effecten; enerzijds tussen *de lengte van de VzC* en *de lengte van het achterveld* en anderzijds tussen *de toegankelijkheid van de referent* en *het informationeel belang van de referent in de nacontext*. Deze interactie-effecten zijn nieuw en nuanceren in het bijzonder de verklaringen in termen van *complexiteit* en *discoursstatus*.

Hoofdstuk 5

Discussie en conclusie

In het vorige hoofdstuk hebben we met behulp van een logistische-regressieanalyse vastgesteld dat een hele resem verklarende variabelen simultaan een effect hebben op de keuze tussen middenveldpositie en achterveldpositie van VzC's. De belangrijkste bevindingen sommen we puntsgewijs op:

- Achterveldpositie is de meest frequente positie voor VzC's in journalistieke teksten (§4.1.1). Dat is een verrassende vaststelling, aangezien zowel traditionele grammatica's als de meer gespecialiseerde literatuur er tot nog toe van uitgingen dat het middenveld de *first choice*-positie is en het achterveld enkel in tweede instantie gekozen wordt (namelijk in het geval dat het middenveld te complex wordt). Daaruit kan men afleiden dat het middenveld de meest frequente positie zou moeten zijn, wat dus niet het geval is voor geschreven journalistiek Belgisch Nederlands.
- De plaatsing van VzC's is een typisch multifactorieel fenomeen, dat in eerste instantie door de semantische status van de VzC bepaald wordt en vervolgens door de complexiteitsvariabelen, de werkwoorden en de discoursvariabelen (§4.2.1 voor een overzicht van de relatieve impact van de variabelen). Dat sommige van die variabelen een invloed hebben op de plaatsing van VzC's wisten we al voordat dit onderzoek uitgevoerd werd, maar met dit onderzoek hebben we een veel vollediger zicht op de variabelen die de plaatsing sturen: we zijn nieuwe variabelen op het spoor gekomen en variabelen waarvan men al lang verwachtte dat ze een effect hebben, zijn weggevallen. We hebben niet alleen een vollediger beeld van de variabelen die de plaatsing van VzC's sturen, we hebben er nu ook een accurater zicht op (in termen van relatieve effecten en *overall*-adequaatheid).
- Er zijn twee interessante interactie-effecten:
 - Het stijgende positieve lineaire effect van de lengte van de VzC op achterveldpositie wordt afgeremd naarmate het achterveld complexer

wordt (§4.2.3.5). Deze vaststelling nuanceert het traditionele idee van het achterveld als een (oeverloos) overloopvat.

- Referenten die in de voorcontext vermeld werden (i.e. toegankelijke referenten) verkiezen in principe middenveldpositie, maar deze tendens wordt afgezwakt naarmate het informationele belang van de referent toeneemt. Omgekeerd verkiezen ontoegankelijke of nieuwe referenten in principe achterveldpositie, maar deze voorkeur wordt afgeremd als de referent veeleer onbelangrijk is (§4.2.3.6).

In de volgende paragrafen staan we uitgebreid stil bij deze bevindingen. In paragraaf 5.1 zullen we eerst de voorkeurspositie van de VzC's in onze corpusdata bespreken. We zullen meer bepaald argumenteren dat in geschreven journalistieke teksten het middenveld niet de meest geprefereerde positie is, in tegenstelling tot wat de ANS (1997) beweert en in tegenstelling tot wat Jansen (1978, 1979) en Braecke (1990) in gesproken taal hebben vastgesteld. Daardoor komt ook het idee van het achterveld als een uitvalsbasis om het middenveld te verlichten op de helling te staan, althans voor geschreven journalistieke teksten.

In paragraaf 5.2 staan we stil bij het multifactoriële karakter van de plaatsing van VzC's in geschreven journalistiek Nederlands. We zullen meer bepaald het effect dat elk van de predictoren heeft, bespreken in volgorde van hun relatieve belang. We beginnen bijgevolg met de semantiek van de VzC en met de complexiteitsvariabelen, en we sluiten af met de werkwoorden en de discoursgerelateerde variabelen.

De bespreking van empirische onderzoeksresultaten is uiteraard nooit theorieneutraal. In wat volgt zullen we het effect van de predictoren niet zozeer kaderen binnen een zeer specifieke linguïstische theorie (zoals de constructiegrammatica), maar binnen veeleer algemene principes die centraal staan binnen zgn. *usage-based theorieën*, principes die hun weg vinden naar allerlei specifieke theorieën. Zo zullen we veronderstellen dat taalsysteem en taalgebruik elkaar wederzijds beïnvloeden. Variatie is bijgevolg inherent aan zowel taalgebruik als taalsysteem:

“While all linguists are likely to agree that grammar is the cognitive organization of language, a usage-based theorist would make the more specific proposal that grammar is the cognitive organization of one's experience with language” (Bybee 2006: 711).

5.1 De VzC's verkiesen achterveldpositie

We hebben in deze studie aangetoond dat het achterveld de meest frequente positie is voor VzC's. Met 57% achterveldpositie wijken onze resultaten af van het eerdere empirische onderzoek van Jansen (1978) en Braecke (1990): Jansen stelde ongeveer 35% achterveldpositie vast en Braecke 40%. Toch spreekt onze bevinding niet noodzakelijk de bestaande studies van Jansen en Braecke tegen, aangezien het genre (geschreven versus gesproken) verschilt. We kunnen echter wel op basis van onze studie concluderen dat vanuit een kwantitatief standpunt het achterveld de geprefereerde positie is voor VzC's in geschreven journalistieke teksten. Dat is op zich zeer relevant voor traditionele grammatica's, zoals de ANS (1997), waarin deze genre-gerelateerde nuance afwezig is.

De voorkeur voor achterveldpositie in geschreven journalistieke teksten doet natuurlijk ook de vraag rijzen waarom sprekers en schrijvers de VzC's anders positioneren in de zin. Er zijn o.i. twee mogelijke, plausibele verklaringen, een corpusgerelateerde en een psycholinguïstische. Voor de corpusgerelateerde verklaring gaan we ervan uit dat de verklarende variabelen die van belang zijn gebleken bij het statistisch modelleren van de geschreven data ook van belang zijn voor gesproken data. Dat is uiteraard een assumptie – er zijn tot op heden geen multifactoriële studies over de plaatsing van VzC's in gesproken Nederlands – maar wel een beredeneerde assumptie: veel van de variabelen die in het vorige hoofdstuk significant bleken te zijn, zijn ook significant voor andere types van woordvolgordevariatie, zowel in het Nederlands als in het Engels (bv. *syntactische complexiteit* in de Engelse Heavy NP shift (Stallings, MacDonald & O'Seaghdha 1998) of in Particle Placement (Hawkins 1994; Gries 2002a); *discursieve toegankelijkheid* in de aan- vs. afwezigheid van *er* (Grondelaers en Speelman 2007); *definietheid* bij het Nederlandse woordvolgordefenomeen scrambling (Van Bergen & de Swart 2010); *semantische klasse* en de *werkwoorden* in de Engelse Dative Alternation (Bresnan et al. 2005)), zowel in geschreven als in gesproken taal (bv. de Heavy NP shift in gesproken taal (Stallings, MacDonald & O'Seaghdha 1998) en in geschreven taal (Wasow 1997)). Onder die assumptie is het interessant vast te stellen dat het aantal korte(re) VzC's in Braeckes onderzoek (naar gesproken Nederlands) relatief groter is dan het aantal in onze dataset: in Braeckes onderzoek telt 94,4% (of 2676/ 2834) van de VzC's 2-8 lettergrepen tegenover 63,3% (of 1092/ 1725) in eigen onderzoek en 63,3% (of 1794/ 2834) telt slechts 2-4 lettergrepen versus 27,9% in eigen onderzoek (481/ 1725). In die zin is het niet verwonderlijk dat in gesproken Nederlands VzC's een grotere voorkeur hebben voor middenveldpositie; ze volgen de functionele tendens om kortere zinselementen meer vooraan de zin te plaatsen, gevolgd door langere zinselementen (cf. het Language-Independent Preferred Order of Constituents van Dik 1978; Hawkins 1994; Wasow 1997; Stallings, MacDonald & O'Seaghdha 1998), net zoals we in het geschreven Nederlands vastgesteld hebben. Samenvattend kunnen we dus zeggen dat we er met deze

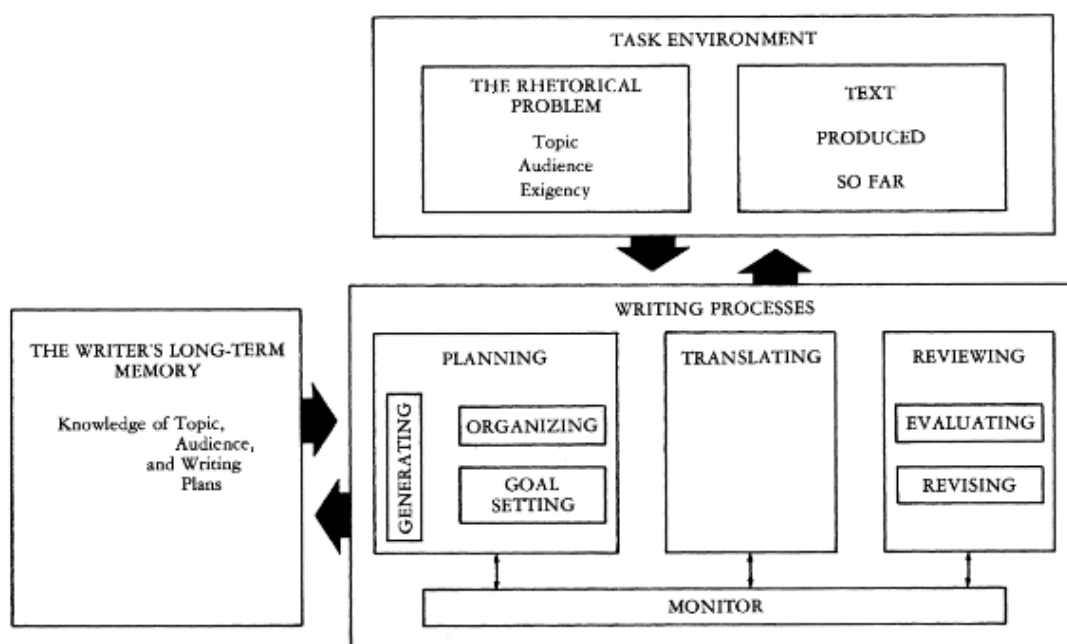
corpusgerelateerde verklaring van uitgaan dat er relatief weinig functionele verschillen zijn in de plaatsing van VzC's tussen gesproken en geschreven Nederlands, i.e. de sturende krachten zijn dezelfde, en het zijn precies die sturende krachten die ervoor zorgen dat VzC's in gesproken Nederlands, die gemiddeld genomen korter zijn dan die in geschreven Nederlands, vaker in het middenveld terechtkomen dan die in geschreven Nederlands. Uiteraard, zoals reeds eerder gezegd: deze verklaring moet in vervolgonderzoek op haar waarde getest worden.

Mocht de corpusgerelateerde verklaring in toekomstig onderzoek verworpen worden, dan lijkt het voor de hand te liggen om het woordvolgordeverschil te relateren aan de verschillende productieomstandigheden. De frequente positionering in het middenveld in gesproken taal versus in het achterveld in geschreven taal zou dan begrepen kunnen worden als een gevolg van de verschillende verwerkingsbehoeften van een spreker (in gesproken omstandigheden) en een lezer (in geschreven omstandigheden). In een spontaan gesproken discours, en zeker tijdens onvoorbereide spontane communicatie, moet de spreker de uiting bedenken, formuleren en online produceren en dat allemaal in een zeer korte tijdsspanne (Levelt 1989). Het idee is daarbij dat sprekers beginnen te spreken zonder de volledige syntactische structuur van de uiting op voorhand te plannen (Levelt 1989; Wheeldon & Lahiri 1997; Ferreira & Swets 2002; Allum & Wheeldon 2007)¹. De basisbepinning is daarbij niet zozeer het interpretatieproces voor de hoorder zo gemakkelijk mogelijk te maken, maar wel alles gezegd te krijgen:

Given the slower time course of writing, there is less need to ensure fast grammatical encoding. A possible implication is that the grammatical encoding processes for speaking and writing diverge; since fluency is a strong requirement for speaking, a fluency-enhancing device such as linearization is necessary. This is not the case for writing. (Hartsuiker & Westenberg 2000: B32).

We kunnen bijgevolg veronderstellen dat de spreker vanuit een spreker-georiënteerd design een uiting opbouwt. Schrijvers, daarentegen, hebben de tijd om op zoek te gaan naar de optimale zinsconstructie voor zichzelf én voor de lezer (*audience-design*). Onderstaande figuur 65 illustreert hoe het cognitieve schrijfproces in zijn werk gaat (Flower & Hayes 1981: 370) en toont aan hoe de schrijver de tekst (meermaals) kan evalueren en reviseren, wat uiteraard minder vanzelfsprekend is in de gesproken taal.

¹ Er zijn twee verschillende visies. Onder anderen Levelt (1989) en Wheeldon & Lahiri (1997) geloven dat sprekers niets op voorhand plannen. Sprekers beginnen te articuleren zodra ze het eerste fonologische woord van hun uiting kennen. Tijdens de formulering plannen ze na het eerste woord het tweede woord, enz. Ferreira & Swets (2002) en Allum & Wheeldon (2007) geloven daarentegen dat spreker niet beginnen te produceren vooraleer ze sommige aspecten van de uiting kunnen memoriseren: 'speakers do not initiate production before having formulated at least some aspects of the utterance that go beyond the immediate phonological word' (Ferreira & Swets 2002: 76).



Figuur 65: Model van het cognitieve schrijfproces volgens Flower & Hayes (1981: 370)

Door het achterveld te verkiezen boven het middenveld, is het werkwoord – het semantisch en syntactisch meest bepalende element voor de zinsinterpretatie – sneller toegankelijk voor de lezer, die als gevolg daarvan een gebalanceerde zin krijgt met zinslelementen voor en na de werkwoordelijke eindgroep.² Deze veronderstelling beantwoordt ook aan de algemene tendens om woorden die gerelateerd zijn, zoals het subject en het werkwoord, dicht bij elkaar te plaatsen (o.a. Hawkins 1994; Gibson 1998, 2000; Gildea & Temperley 2010). Zeker in een geschreven context, waar tijd geen beperking is, lijkt deze verklaring zeer valabel: “it seems more likely, rather, that syntactic choices in writing reflect strategies to facilitate comprehension” (Temperley 2007: 328). Schrijvers hebben bijgevolg de tijd en de mogelijkheid om vanuit een lezer-georiënteerd design te werken.

Samengevat toont deze studie aan dat het middenveld niet de standaardpositie is voor VzC’s in geschreven journalistieke teksten, wat we op basis van de bestaande grammatica’s en het bestaande empirische onderzoek wel zouden verwachten. De theoretische conclusie die wij aan onze observatie kunnen verbinden, is dat het structuralistische idee van het middenveld als basispositie en het achterveld als een uitvalsbasis verlaten moet worden ten gunste van een theoretisch model waarbij het

² Om na te gaan hoe zo’n evenwichtige zin er precies uitziet, hebben we het gemiddelde berekend van het aantal woorden in het middenveld (plus de VzC als de VzC in het middenveld staat) en het gemiddelde voor het aantal woorden in het achterveld (plus de VzC als de VzC in het achterveld staat). Het resultaat is dat de gemiddelde zin 4,14 woorden in het middenveld telt en 5,30 woorden in het achterveld, een min of meer gebalanceerde verdeling wat aan onze verwachting beantwoordt.

middenveld en het achterveld als twee equivalente posities worden beschouwd, althans voor het geschreven Nederlands in journalistieke teksten. In het geschreven discours is het middenveld geen basispositie, maar zijn het middenveld en het achterveld twee posities die in beide richtingen linguïstisch gevuld kunnen worden. Schrijvers kunnen nu eenmaal, in hun zoektocht naar een evenwichtige verdeling, zinsdelen uit het middenveld naar het achterveld verschuiven, maar evengoed vanuit het achterveld naar het middenveld. Deze hypothese spreekt natuurlijk ook de eenrichtingsbeweging in de traditionele literatuur tegen, waarbij de VzC vanuit het middenveld naar het achterveld verplaatst wordt, en spreekt in het voordeel van een bidirectionele beweging. Deze hypothese moet natuurlijk in de toekomst met behulp van een experimenteel onderzoek verder uitgewerkt en getest worden.

5.2 De predictoren op een functioneel-cognitieve manier verklaard

5.2.1 De semantische status van de VzC

De variabele die de grootste impact heeft op de positie van VzC's is de semantische status van de VzC. Tijds- en plaatsbepalingen³ staan significant vaker in middenveldpositie in vergelijking met semantische klassen zoals *beperking*, *doel*, *resultaat*, enz. Tegelijkertijd hebben we ook een verschil vastgesteld tussen tijds- en plaatsbepalingen. Het resultaat is een schaal van zeer weinig kans op achterveldpositie (*tijd*), evenveel kans op achterveldpositie als middenveldpositie (*plaats* en de andere semantische groepen die deel uitmaken van categorie 2) en zeer veel kans op achterveldpositie (*beperking* en de andere semantische groepen die deel uitmaken van categorie 1 en de voorzetselvoorwerpen).

Een mogelijke verklaring waarom net tijds- en plaatsbepalingen significant vaker in het middenveld staan, is dat ze enkel het gezegde situeren en niet beschrijven. Bepalingen van *tijd* en *plaats* kaderen de handeling binnen een bepaalde periode of een bepaalde situatie waarin zich iets afspeelt. Ze geven dus enkel een context aan waarbinnen het gezegde begrepen moet worden. Daartegenover staan bepalingen van *beperking*, *doel*, *resultaat*, enz. die de beperking, het doel en het resultaat van de handeling definiëren. Deze laatste groep beschrijft bijgevolg het gezegde en staat daarom ook semantisch dicht bij de handeling an sich. Eenzelfde redenering vinden we terug bij Quirk et al. (1985). Volgens Quirk et al. hebben SVO-talen zoals het Engels een voorkeur voor een MPT-regel (*manner* < *place* < *time*), en stellen we het tegenovergestelde vast voor SOV-talen zoals het Nederlands (dus *tijd* < *plaats* < *wijze*). Quirk et al. stellen dat een bepaling van *wijze* semantisch meer verbonden is met het gezegde dan tijds- en plaatsbepalingen⁴. Daardoor verkiezen ze ook een positie dicht bij de werkwoordelijke eindgroep. Daarnaast zijn plaatsbepalingen meer dan tijdsbepalingen, maar minder dan wijzebepalingen, verbonden met het gezegde.

De volgorde *tijd* < *plaats* < *andere bepalingen* vinden we ook in de ANS (1997) terug en in het onderzoek van Braecke (1990). Hoewel de ANS (1997) de semantische status van de VzC niet vermeldt als een factor die de woordvolgorde van VzC's bepaalt, veronderstelt

³ Naast *omstandigheid*, *hoedanigheid*, *middel* en *maat* die ook tot categorie 2 behoren maar in mindere mate voorkomen

⁴ We merken hierbij wel op dat beperkingen van *middel* en *hoedanigheid*, ook wel wijzebepalingen, niet significant verschillen van plaatsbepalingen wat hun positie in de zin betreft. Ze maken bijgevolg geen deel uit van categorie 1 maar wel van categorie 2 (zie ook §3.2.3).

ze wel dat als er meerdere bijwoordelijke bepalingen in het middenveld optreden, de volgorde *tijd < plaats < andere bepalingen* geldt. Deze functionele regel kan bijgevolg ook de positie van VzC's in de zin beïnvloeden, maar verklaart op zich niet waarom tijdsbepalingen voor plaatsbepalingen staan en waarom de andere bepalingen hierachter komen te staan.

Naast een semantische interpretatie kan de volgorde van de VzC's ook een discursieve rol spelen. Grondelaers et al. (2009: 154) zien in hun studie naar presentatief *er* dat plaatsbepalingen de andere constituenten in de zin van een context voorzien, en bij uitstek de constituenten die nieuwe informatie bevatten. Plaatsbepalingen (of locatieve bepalingen) scheppen met andere woorden een kader waarbinnen constituenten met nieuwe(re) informatie geïntroduceerd kunnen worden:

In locative inversion-constructions [...], thematic, syntactic, semantic, and real-world properties of an upcoming new entity are made available by the initial adjunct and verb. [...] In Grondelaers et al. (2002), corpus evidence was cited which showed that owing to this “high-constraint” syntax and semantics, locative inversion structures are used significantly more frequently than other constructions for the introduction of new information into the discourse. (Grondelaers et al. 2009: 154)

Wat Grondelaers et al. beweren over plaatsbepalingen, kan o.i. in sommige gevallen ook voor tijdsbepalingen gelden (er wordt bijgevolg een tijds kader gecreëerd waarbinnen de nieuwe informatie begrepen moet worden). We onderzochten daarom in onze dataset hoeveel tijds- en plaatsbepalingen voor het subject staan, i.e. de positie direct na het onderschikkend voegwoord. Deze positie is interessant omdat ze zo goed als altijd gereserveerd is voor discursief toegankelijke informatie, meestal het subject van de zin, of voor constituenten die op zich niet toegankelijk zijn, maar die andere constituenten in de zin contextualiseren en daardoor toegankelijker maken. We stellen vast dat 19% van de tijdsbepalingen voor het subject staat; 12,5% van de plaatsbepalingen en slechts 6,07% van de bepalingen van *beperking*, *doel*, *resultaat*, enz.

Op basis van de positie in de zin kunnen we dus vaststellen dat tijdsbepalingen (en in iets mindere mate plaatsbepalingen) in vergelijking met de andere semantische categorieën meer in de gemarkeerde positie voor het subject staan. Dat laatste kan ook wijzen op een lossere band met het werkwoord. Zoals reeds besproken in 3.2.3 is de variabele *semantische status van de VzC* een combinatie van een puur grammaticale onderverdeling en een semantische. Het is dus mogelijk dat een tijdsbepaling semantisch het minst verbonden is met het werkwoord, gevolgd door een plaatsbepaling, enz. Het zou ook kunnen verklaren waarom er geen significant verschil is tussen de semantisch beschrijvende categorieën (met *beperking* als de meest geattesteerde semantische categorie) en de fixed-categorie (of de voorzetselvoorwerpen/-objecten). Beide categorieën zijn sterk verbonden met het werkwoord (semantisch of grammaticaal). In

combinatie met de Nederlandse zinsstructuur, waarbij de positie voor de werkwoordelijke eindgroep ingenomen kan worden door een inherent zinsdeel, kan dat tot een voorkeur voor achterveldpositie leiden.

De positie van de semantische categorieën kan ook de voorspelbaarheid van de overige zinsdelen in de zin, of van de zin in het algemeen, laten toenemen. Eye-tracking studies tonen namelijk aan dat een taalgebruiker voorspellingen maakt op basis van eerdere informatie in de zin (o.a., Altmann & Kamide 1999; Kamide et al. 2003). Volgens Kamide et al. (2003) zal de taalgebruiker bij elk stukje nieuwe informatie voorspellingen maken over de grammaticale rol van de gelezen (of gehoorde) zinsdelen, maar ook voorspellen welke informatie er nog aan bod moet komen.

To interpret a sentence incrementally, word-by-word, requires the partial interpretation of what has been encountered thus far—thus, after *The boy will eat*, some representation must be constructed which at the very least places *eat* and *the boy* in a subject–predicate relationship. Beyond that, it may also reflect the semantic properties of *boy*, and of *eat* (including that the event being referred to takes place in the future). And given the semantics of *eat*, a partial interpretation may also reflect that fact that the boy will eat *something*. But the fullest possible interpretation would include also the interpretation at each moment in time of the consequences of what has been encountered so far for what may come next (both in the language and in the mental or real world which that unfolding language describes). (Kamide et al. 2003)

We illustreren het voorgaande aan de hand van enkele voorbeelden uit de eigen dataset. In zin 102 krijgt de lezer eerst het subject te lezen, *de Verenigde Staten*, en vervolgens de VzC van *tijd*, namelijk *op 11 september 2011*. Deze tijdsbepaling in combinatie met de *Verenigde Staten* roept automatisch *terroristische aanslagen* op nog voor de lezer de VzC in achterveldpositie heeft gelezen. De VzC in het achterveld, en zelfs het gezegde (*slachtoffer zijn geworden*) zijn dus heel voorspelbaar op basis van het tijds kader dat hier wordt gecreëerd.

- (102) Ja, ik weet wel <begin>dat <su>de Verenigde Staten</su> <vz>op 11 september 2001</vz> het slachtoffer <v>zijn geworden</v> van terroristische aanslagen <end> en dat sommige van die terroristen de Verenigde Staten zijn binnengekomen met J-1 visa.

In zin 103 kan de lezer (na het lezen van *het veld*) wel voorspellen dat supporters van een bepaalde voetbalclub *het veld zullen oplopen of bestormen* en dat ze daarbij ofwel zeer vrolijk zullen zijn (bij winst) of ontzet (bij verlies).

- (103) <begin>Dat <su>tientallen aanhangers van zijn eigen club</su> <vz>tijdens het duel tegen Vitesse</vz> het veld <v>bestormden</v> om hun ongenoegen duidelijk te maken <end>, was voor Frans Adelaar (45) de aanleiding om vorige week maandag zijn vertrek aan te kondigen.

In zin 104 weet de lezer na de tijdsbepaling dat *men iets wil vestigen/oprichten/bereiken*. Na het lezen van het direct object neemt de voorspelbaarheid van het werkwoord alleen maar toe.

- (104) Hoewel de ontwikkeling van de Amerikaanse markt geen prioriteit is, hopen we toch <begin>dat <su>we</su> <vz>over vijf jaar</vz> een masterfranchise <v>kunnen vestigen</v> <end>.

Hetzelfde geldt voor zin 105 met een plaatsbepaling in het middenveld. Nadat de lezer de plaatsbepaling *in Shanghai* heeft gelezen kan je al voorspellen dat *iets ergens plaatsvindt of gebeurt*.

- (105) Conclusie: ja, het is meer dan terecht <begin>dat <su>de aftrap van Live Earth</su> <vz>in Shanghai</vz> <v>plaatsvindt</v> <end>.

We veronderstellen dus dat de positie van de VzC ook cognitieve voordelen biedt en dat het de voorspelling van nieuw materiaal kan bevorderen.

Tot nu toe hebben we de *fixed*-groep bij de bespreking van deze variabele grotendeels links laten liggen. Deze *fixed*-groep valt samen met de grammaticale functie *voorzetselvoorwerp* en omdat een voorzetselvoorwerp opgeroepen wordt door een werkwoord of door de combinatie van een werkwoord met een object, lijkt het logischer om het effect ervan te verbinden met de werkwoorden. Het effect van de voorzetselvoorwerpen – voorzetselvoorwerpen stimuleren namelijk achterveldpositie – zal bijgevolg ook in paragraaf 5.2.3 ter sprake komen.

5.2.2 De syntactische complexiteit

Voor de *syntactische complexiteit* beperkten we het onderzoek niet enkel tot de complexiteit van de VzC, maar berekenden we ook de complexiteit van het middenveld (tussen de kern van het subject en het vervoegde werkwoord) en de complexiteit van het achterveld. Daarnaast operationaliseerden we *complexiteit* op drie manieren en telden we achtereenvolgens 1. het aantal woorden, 2. het aantal lettergrepen en 3. het aantal inbeddingen. Uit onze studie bleek dat de verschillende operationalisaties correleerden; lengte in woorden correleerde met lengte in lettergrepen en lengte in woorden met het aantal hiërarchische inbeddingen (§4.1.2.4). We namen daarom slechts één operationalisatie mee in het statistische model: de lengte in woorden.

De logistische regressie toonde aan dat de drie variabelen een rol spelen, maar weerlegt het idee dat de *complexiteit* (in de literatuur ook wel *grammatical weight*) de allesbepalende factor is die de woordvolgordevariatie definieert (in tegenstelling tot Hawkins 1992, 1994). In deze paragraaf zullen we de verschillende complexiteitsvariabelen bespreken. We beginnen met het hoofdeffect de *lengte van het*

middenveld en bespreken vervolgens het interactie-effect van de *lengte van de VzC* en de *lengte van het achterveld*. Een individuele beschrijving van de *lengte van de VzC* en de *lengte van het achterveld* is niet relevant, net omdat deze variabelen interageren en ze door elkaar beïnvloed en bepaald worden.

Met behulp van de multivariate analyse hebben we vastgesteld dat de lengte van het middenveld een positief lineair effect heeft op de positie van VzC's, ook wanneer er tegelijkertijd rekening gehouden wordt met de invloed van andere relevante predictoren. De kans op achteropplaatsing stijgt met andere woorden naarmate de lengte van het middenveld stijgt. Dit resultaat kunnen we ook verklaren aan de hand van onze eerdere bevinding in paragraaf 5.1, namelijk dat schrijvers vanuit een lezer-georiënteerd design werken en op zoek gaan naar een min of meer gebalanceerde zin. Daardoor is ook het werkwoord (de tweede pool én het structurerende element van de zin) sneller zichtbaar en is bijgevolg de overkoepelende structuur (nl. de relatie tussen de twee gerelateerde zinsdelen: subject en persoonsvorm) sneller toegankelijk. Dat laatste komt volgens de literatuur ook de interpreteerbaarheid ten goede (o.a. Van Haeringen 1947; ANS 1997; Hawkins 1994; Gibson 2000).

Hoewel het effect overeenkomt met de verwachting om de twee polen niet te ver van elkaar te plaatsen, moeten we toch stilstaan bij twee deelaspecten, met name de vaststelling dat lege middenvelden (geen woorden tussen subject en persoonsvorm) niet opvallend vaker een VzC aantrekken dan een achterveld, en dat VzC's in het geval van zeer lange middenvelden geen absolute voorkeur hebben voor achterveldpositie. Hetzelfde resultaat werd trouwens ook door Jansen (1978, 1979) voor gesproken taal vastgesteld. Als het de bedoeling is van de taalgebruiker (spreker of schrijver) om de cognitieve verwerkingslast te verminderen voor de toehoorder door de overkoepelende structuur zichtbaar en toegankelijk te maken, waarom is de kans op achterveldpositie voor de VzC's dan niet absoluut?

Om deze laatste vraag te beantwoorden, verwijzen we naar het onderzoek van Gildea & Temperley (2010). Zij onderzochten de *Dependency Length Minimization*-theorie (cf. Gibson 1998, 2000), die stelt dat talen een voorkeur hebben om woorden die gerelateerd zijn dicht bij elkaar te plaatsen ("preference for closely related words to be close together in the sentence"; Gildea & Temperley 2010: 286), en die dus vergeleken kan worden met de *verlicht-de-tang-regel* in het Nederlands. Om na te gaan of talen rekening houden met de *Dependency Length Minimization* (ook wel DLM) onderzochten Gildea & Temperley (2010) o.a. de afstand tussen het subject en het vervoegde werkwoord in het Engels en in het Duits. De auteurs vertrekken vanuit het idee dat de DLM-regel sterker aanwezig zou zijn in het Duits dan in het Engels, net omdat de Duitse woordvolgorde (net zoals de

Nederlandse) vrijer is dan het Engels⁵. Het Duits zou dus meer in staat zijn om te streven naar een *Dependency Length Minimization*, wat het Engels met zijn meer starre woordvolgorde niet altijd kan.

German offers an interesting comparison with English as it has somewhat more freedom of word order; its word order is sometimes described as being “semi-free” or free within certain constraints (Kempen & Harbusch, 2008; Webelhuth, 1992). Thus, in a language with free (or semi-free) word order, one might expect to see greater evidence of DLM. (Gildea & Temperley 2010: 297)

Gildea & Temperley (2010: 300) stelden echter vast dat de *Dependency Length Minimization* minder aanwezig is in het Duits in vergelijking met het Engels: “the effect of DLM is not as strong in German as it is in English”. De verklaring is volgens hen de verplichte rechtse positie van het vervoegde werkwoord in o.a. bijzinnen.

Of particular importance is the fact that verbs in German clauses are often clause-final (all participles, as well as finite verbs in dependent clauses). A participle may be at great distance from the preceding auxiliary verb, with NPs, PPs, and other things in between; in English, by contrast, auxiliaries and participles are almost always adjacent. In a dependent clause, similarly, the clause-final head verb of the clause may be far from its head, which most often precedes the clause—for example, the relative pronoun of a relative clause. The placement of the verb at the end of the clause also means that a crowding effect may occur, with several dependent phrases all on the same side; in English, by contrast, one dependent—the subject NP—normally branches to the left, while others branch to the right, creating a more “balanced” configuration. (Gildea & Temperley 2010: 300)

Wat de woordvolgorde betreft, bekleedt het Nederlands een tussenpositie, tussen het Engels en het Duits. In tegenstelling tot bijvoorbeeld het (Hoog)Duits (Van de Velde 1973: 141) staat het Nederlands bekend om zijn flexibiliteit om elementen achter de werkwoordelijke eindgroep te plaatsen. Anderzijds moet het werkwoord (of de werkwoorden) verplicht achteraan staan in een bijzin – zij het in minder extreme mate dan in het Duits (De Sutter & Van de Velde 2008) – en wordt er automatisch een zekere afstand gecreëerd tussen subject en werkwoord. De Nederlandstalige taalgebruiker kan door de vaste positie van het werkwoord niet altijd streven naar de ‘ideale’ woordvolgorde volgens de DLM-theorie (i.e. de woordvolgorde waarbij de woorden die bij elkaar horen, zo dicht mogelijk bij elkaar staan). We kunnen onszelf dan ook de vraag stellen of deze ‘ideale’ volgorde volgens de DLM-theorie wel de ideale volgorde in het

⁵ Om de vrijere woordvolgorde te illustreren, verwijzen Gildea & Temperley naar de volgorde van objecten in de zin: “A well-known example of this freedom is the ordering of arguments of the verb (subject, object, and indirect object), which is highly variable in German (Heylen, 2005; Kempen & Harbusch, 2008)” (Gildea & Temperley 2010: 297).

Nederlands (en in het Duits) is. Gildea & Temperley (2010: 301) stelden namelijk in hun onderzoek vast dat de gemiddelde afstand tussen subject en vervoegde werkwoord in het Duitse corpus gelijk is aan 3,06 woorden. Berekenen we de gemiddelde afstand tussen het subject en het werkwoord in onze dataset (plus de VzC als de VzC tussen het subject en het werkwoord staat), dan krijgen we een resultaat van 4,67 woorden. Bijgevolg zouden we evengoed kunnen veronderstellen dat een middenveld van ongeveer 3 à 4 woorden (bijvoorbeeld met een relatief korte VzC in het middenveld) ideaal is. Het zou ook verklaren waarom zinnen met een leeg middenveld evenveel kans hebben op middenveldpositie als op achterveldpositie, aangezien de taalgebruiker o.i. niet per se naar een leeg middenveld streeft. Daarnaast stelden we ook vast dat VzC's in het geval van een lang middenveld geen absolute voorkeur voor achterveldpositie hebben. De verklaring zou kunnen zijn dat de aanwezigheid van andere zinselementen in het middenveld, het werkwoord zo voorspelbaar maakt (zie ook §5.2.1), dat een opbouw waarbij het subject en het werkwoord niet te ver van elkaar geplaatst worden (en dus met een korter middenveld en een VzC in het achterveld) overbodig is. Bovendien heeft ook het geschreven medium hier een voordeel, want het biedt de lezer de mogelijkheid om de zin te herlezen bij eventuele moeilijkheden om de overkoepelende structuur te verwerken.

Kort samengevat tonen de resultaten van de variabele *lengte van het middenveld* aan dat schrijvers (en ook sprekers in navolging van Jansen 1978, 1979) wel rekening houden met de cognitieve verwerking van een lang middenveld, en dus ook rekening houden met de cognitieve last voor een toehoorder, maar niet in die mate dat ze subject en werkwoord onmiddellijk naast elkaar plaatsen.

We stelden in deze studie ook vast dat de *lengte van de VzC* en de *lengte van het achterveld* met elkaar interageren. Dat betekent concreet dat het sterke positieve lineaire effect van de *lengte van de VzC* op achterveldpositie afgeremd wordt als de lengte van het achterveld toeneemt. Dat de *lengte van de VzC* een effect heeft, werd al in eerder empirisch onderzoek vastgesteld. De complexiteit van de VzC bleek zelfs een constante te zijn in het onderzoek naar dit type van woordvolgordevariatie (Jansen 1978, 1979; De Meersman 1980; Braecke 1990; Hoeksema 2014). Bovendien past het resultaat van de lengte van de VzC ook binnen de cross-linguïstische, functionele tendens om langere constituenten achteraan te plaatsen, ook wel bekend als *end-weight* (Hawkins 1994; Wasow 1997; Stallings, MacDonald & O'Seaghdha 1998).

Het effect van de *lengte van de VzC* past ook binnen de DLM-theorie, want een lange VzC in het middenveld kan de afstand tussen twee gerelateerde zinsdelen, in dit geval het subject en het werkwoord, vergroten ("a long complement will greatly lengthen the dependency from the clause-final verb to its preceding head"; Gildea & Temperley 2010: 298). Om de afstand enigszins te beperken (tot bijvoorbeeld 3 à 4 woorden), wordt de lange VzC geëxtraponeerd. Tegelijkertijd komt het effect van de lengte van de VzC ook overeen met wat we verwachten op basis van de *Early Immediate Constituent*-theorie van Hawkins

(1994), wat later onder het *Minimize Phrasal Domain* van het *Minimize Domains*-principe valt (Hawkins 2004). Deze theorie kunnen we trouwens vergelijken met de DLM-theorie en stelt dat het verwerkingsproces van de zin afhangt van de afstand tussen twee (of meerdere) gerelateerde woorden (voor een vergelijking tussen deze twee theorieën, zie Temperley 2007). Hawkins beschouwt de woordvolgorde als een manier om de syntactische structuur van een zin zo efficiënt mogelijk te analyseren. Daarbij hoeft de taalgebruiker niet te wachten tot het laatste woord van de zin. Zodra de taalgebruiker het hoofdwerkwoord heeft bereikt en zodra hij voldoende informatie heeft over hoe de laatste constituent is opgebouwd, kan de structuur van de laatste constituent en de hele zin bepaald worden. In het geval er twee of meer woordvolgordes mogelijk zijn, krijgt die volgorde de voorkeur die de taalgebruiker het snelst in staat stelt om de zinsstructuur te bepalen.

Ter illustratie: in onderstaande voorbeeldzin uit onze dataset kan de VzC [_{VzC}voor de managementcapaciteiten die ze bezit] in het achterveld (zin 106a) of in het middenveld staan. In het middenveld staat het ofwel voor het gezegde (106b), ofwel voor het subject (106c):

- (106) a ...dat [_{NC}het bedrijf] [_{AdjCte} groot] [_{VP}was geworden] [_{VzC}voor de managementcapaciteiten die ze bezit].
 b ...dat [_{NC}het bedrijf] [_{VzC}voor de managementcapaciteiten die ze bezit] [_{AdjCte} groot] [_{VP}was geworden].
 c ...dat [_{VzC}voor de managementcapaciteiten die ze bezit]. [_{NC}het bedrijf] [_{AdjCte} groot] [_{VP}was geworden]

Berekenen we vervolgens de afstand van alle zinsdelen in de bijzin (Hawkins spreekt over *immediate constituents* of *ICs*) tot het hoofd (in dit geval het koppelwerkwoord *worden*) dan stellen we vast dat in zin 106a de zinsdelen sneller herkend worden. In zinnen 106b en c moet je alle zinsdelen lezen tot het hoofdwerkwoord *worden*.

a	dat	[_{NC} het bedrijf]		[_{AdjC} te groot]		[_{VP} was geworden]		[_{VzC} voor de managementcapaciteiten die ze bezit]					
	0	1	2	3	4	5	6	7					
b	dat	[_{NC} het bedrijf]		[_{VzC} voor de managementcapaciteiten die ze bezit]					[_{AdjC} te groot]		[_{VP} was geworden]		
	0	1	2	3	4	5	6	7	8	9	10	11	12
c	dat	[_{VzC} voor de managementcapaciteiten die ze bezit]					[_{NC} het bedrijf]		[_{AdjC} te groot]		[_{VP} was geworden]		
	0	1	2	3	4	5	6	7	8	9	10	11	12

Volgens Hawkins zal de volgorde die de minste verwerking vraagt en die bijgevolg efficiënter is, de voorkeur krijgen. Deze efficiëntie kan ook berekend worden door het aantal ICs (in totaal zijn er 4 zinsdelen) te delen door het aantal verbindingen die nodig zijn om het hoofdwerkwoord met het (begin van het) zinsdeel te verbinden:

- a aantal ICs/ aantal verbindingen: $4/7 = 57\%$
 b aantal ICs/ aantal verbindingen $4/12 = 33\%$
 c aantal ICs/ aantal verbindingen $4/12 = 33\%$

Hoe hoger het percentage, hoe efficiënter de woordvolgorde, aangezien de taalgebruiker minder verbindingen nodig heeft en toch dezelfde informatie over de zinsdelen (het aantal en de vorm) krijgt. Zin 106a met een VzC in het achterveld is bijgevolg de meest efficiënte woordvolgorde.

Onderhavige studie toont aan dat de taalgebruiker in een geschreven tekst rekening houdt met de complexiteit van de VzC, maar dat er tegelijkertijd ook rekening wordt gehouden met wat er (voor én) na de werkwoordelijke eindgroep staat. De meest efficiënte volgorde is dus niet altijd de volgorde met een VzC in het achterveld, zeker niet als er bijvoorbeeld een bijzin, een andere VzC, enz. in het achterveld staat. De meest efficiënte constructie blijkt namelijk een constructie te zijn waarbij alle velden en de VzC een rol spelen, met een gebalanceerde verdeling van het taalmateriaal over het middenveld en het achterveld als resultaat. Volgens Gildea & Temperley is de volgorde in het Engels meer in balans, omdat het subject links van het werkwoord staat en de andere objecten rechts. Berekenen we in onze dataset het aantal woorden in het middenveld (tussen subject en werkwoord, plus de VzC als de VzC tussen het subject en het werkwoord staat) en in het achterveld, dan zien we eenzelfde gebalanceerd resultaat in de Nederlandstalige journalistieke teksten van gemiddeld 4 woorden in het middenveld en 5 woorden in het achterveld.

We kunnen besluiten dat de *lengte van de VzC* inderdaad een belangrijke variabele is, maar dat deze variabele tegelijkertijd geremd wordt door een andere variabele, m.n. de *lengte van het achterveld*. Het effect van de VzC is dus minder rechtdoorzee dan eerdere studies doen vermoeden. Vanuit een gesproken standpunt is het niet vreemd dat de lengte van het achterveld geen aandacht kreeg. Het zou namelijk betekenen dat sprekers op het moment dat ze een uiting formuleren ook denken aan wat er al na het werkwoord, en dus in het achterveld, staat. Het lijkt logischer dat sprekers zich bezighouden met wat ze moeten onthouden en dat ze oplopend de zinselementen plannen. Het achterveld wordt dan de uitvalsbasis voor de constituenten die de spreker nog niet gezegd heeft in de tangconstructie. In een geschreven stuk hebben schrijvers echter wel de mogelijkheid om langer na te denken over de woordvolgorde in de zin en in de rest van de tekst én om die woordvolgorde nog te wijzigen. Het achterveld kan bijgevolg evengoed de positie van VzC's beïnvloeden, wat ook het geval is in geschreven journalistieke taal.

5.2.3 De werkwoordsvariabelen: type hoofdwerkwoord en zinsrelator

Om de invloed van het werkwoord te operationaliseren, onderzochten we het effect van de *frequentie van het werkwoord*, de *semantische classificatie van de werkwoorden*, het *type hoofdwerkwoord* en de *opbouw van de zinsrelator* (zie ook §3.2.4). Enkel de laatste twee operationalisaties bleken een significant effect te hebben op de positie van VzC's in de Nederlandse journalistieke bijzin. We stelden daarbij vast dat koppelwerkwoorden

significant vaker achterveldpositie verkiezen in vergelijking met zelfstandige werkwoorden en dat samengestelde werkwoorden (werkwoordelijke uitdrukkingen of samengestelde predicaten), reflexieve werkwoorden, partikelwerkwoorden of werkwoorden bestaande uit een combinatie van de voorgaande werkwoorden, een voorkeur hebben voor een VzC in het achterveld in vergelijking met enkelvoudige werkwoorden. Aangezien het effect van het werkwoord nog niet eerder onderzocht werd als een verklaring voor de woordvolgordevariatie van VzC's en aangezien het werkwoord wel degelijk een significant effect blijkt te hebben, zorgt dit onderzoek voor een accurater en vollediger beeld waarbinnen we de woordvolgordevariatie van VzC's moeten begrijpen.

In een volgende stap willen we verklaren waarom net bovengenoemde werkwoorden een VzC in achterveldpositie verkiezen. Het valt op dat de werkwoorden telkens gecombineerd worden met een ander zinselement, gaande van een gezegde, het niet-werkwoordelijke deel van de uitdrukking, een reflexief voornaamwoord of een partikel. Bij alle vermelde werkwoorden is er dus de aanwezigheid van een verplicht, inherent zinselement in het middenveld. Bovendien is de positie van de VzC bij de meeste van deze werkwoorden beperkt. Het gezegde, het niet-werkwoordelijke deel van de uitdrukking en het partikel staan namelijk altijd achteraan het middenveld, net voor de werkwoordelijke eindgroep. Daartussen kan de VzC niet voorkomen (i.e. het inherentieprincipe, ANS 1997). De VzC kan wel voor het inherente zinselement in het middenveld staan (*VzC < inherent element < werkwoord*) of na de werkwoordelijke eindgroep in het achterveld (*inherent element < werkwoord < VzC*). Onze resultaten tonen aan dat taalgebruikers ervoor opteren om de VzC in het achterveld te plaatsen.

Een eerste verklaring is de individuele voorkeur van deze werkwoorden (cf. Ellis 2002a; Holmes, Stowe & Cupples 1989; Kidd, Lieven & Tomasello 2006; Tily et al. 2008). Tily et al. (2008) stelden namelijk met behulp van een eye-tracking studie vast dat taalgebruikers kennis hebben over welke werkwoorden een welbepaalde volgorde verkiezen.⁶ Deze kennis zorgt er ook voor dat de taalgebruiker voorspellingen kan maken over wat er nog gezegd zal worden in de tekst, wat vanzelfsprekend de tekstverwerking vereenvoudigt. Het is dus goed mogelijk dat koppelwerkwoorden, samengestelde werkwoorden, reflexieve werkwoorden en partikelwerkwoorden meer achterveldpositie verkiezen in vergelijking met zelfstandige werkwoorden en met enkelvoudige werkwoorden. Het zou ook kunnen verklaren waarom sommige zelfstandige werkwoorden en enkelvoudige werkwoorden toch achterveldpositie verkiezen boven middenveldpositie. Elk werkwoord vertoont een eigen patroon en heeft een (individuele) voorkeur voor een VzC in het

⁶ Ze onderzochten in concreto de datiefalternantie. Bepaalde werkwoorden tonen daarbij een duidelijke voorkeur voor een dubbelobject-constructie, andere werkwoorden voor de constructie met een VzC (de zogenaamde aan-constructie).

midden- en/of achterveld. Het heeft echter geen zin om de verschillende individuele werkwoorden als een vaste variabele op te nemen in een logistische regressie, aangezien een variabele met 571 levels of waarden “terribly ‘dominant’ and terribly uninterpretable” zou zijn (Speelman 2014: 531). We zouden wel de individuele werkwoorden als een random factor in een *Generalized Linear Mixed Model* kunnen opnemen. We hebben er echter in dit onderzoek voor geopteerd om de werkwoorden in vaste variabelen om te zetten, om zo meer vat te krijgen op welk type werkwoord meer of minder gecombineerd wordt met een VzC in het midden- en/of achterveld.

Een tweede verklaring is de betekenis van het werkwoord en de (mentale) verwerking van het werkwoord (of de uitdrukking) door de taalgebruiker. We hebben namelijk met behulp van de *Collostructional Analysis* van Gries en Stefanowitsch (2004) vastgesteld dat voornamelijk semantisch lege werkwoorden – werkwoorden zoals *hebben* en *zijn* die heel vaak deel uitmaken van uitdrukkingen en samengestelde predicaten – significant vaker achterveldpositie verkiezen. Net deze uitdrukkingen en samengestelde predicaten zijn interessant als het gaat over hoe ze precies opgeslagen zijn in het geheugen. Uitdrukkingen, bijvoorbeeld, zijn verbonden met de letterlijke betekenis van de woorden in de uitdrukking (zoals de betekenis van *rol*, bijvoorbeeld een theaterrol, en *spelen*, zoals optreden in een theaterstuk), maar evengoed met de figuurlijke betekenis van de uitdrukking *ergens een rol in spelen* (namelijk *belangrijk zijn, invloed hebben op iets*).

Idioms are conventionalized word sequences that usually contain ordinary words and predictable morphosyntax, but have extended meaning (usually of a metaphorical nature), as in these examples: *pull strings, level playing field, too many irons in the fire*. Idioms are acknowledged to need lexical representation because of the unpredictable aspects of their meaning, but as Nunberg and colleagues (1994) point out, they are not completely isolated from related words and constructions since many aspects of their meaning and form derive from more general constructions and the meaning of the component words in other contexts. Idioms provide evidence for organized storage in which sequences of words can have lexical representation while still being associated with other occurrences of the same words. (Bybee 2006: 713).

Naast *idiomatische uitdrukkingen* heeft Bybee (2006: 713) het ook over *prefabs*: “Prefabs are word sequences that are conventionalized, but predictable in other ways”. Voorbeelden van *prefabs* zijn partikelwerkwoorden (*finish up, burn down*; Bybee 2006: 713) en werkwoorden die een vast voorzetsel oproepen (*interested in, think of, think about*; Bybee 2006: 713). Onderzoek van Erman & Warren (2000) heeft vastgesteld dat ook deze *prefabs* op een welbepaalde manier opgeslagen zitten in het geheugen:

Speakers recognize prefabs as familiar, which indicates that these sequences of words are stored in memory despite being largely predictable in form and meaning. (Bybee 2006: 713)

Omdat deze *uitdrukkingen* en *prefabs*, ook wel *multiwords*⁷, zeer frequent gebruikt worden, zijn ze als een eenheid opgeslagen, waardoor ze ook makkelijker te verwerken en te produceren zijn, en zorgen voor vooraf gespecificeerde lexicale keuzes (en dus ook woordvolgordes):

Given the high use of prefabs and idioms in natural speech, it appears that a good deal of production (and perception) refers to sequences of prespecified lexical choices rather than to abstract grammar. For this reason, a model that builds a grammar from specific instances of language use, such as an exemplar model or a connectionist model, seems appropriate. (Bybee 2006: 716)

Vertalen we dat naar onze studie, dan kunnen we veronderstellen dat de combinatie van *inherent element* < *werkwoord* < VzC als een eenheid wordt geactiveerd en bijgevolg makkelijker wordt geformuleerd en verwerkt. Door deze vaste opgeslagen structuur denkt de taalgebruiker niet na over een andere mogelijke woordvolgorde, namelijk VzC < *inherent element* < *werkwoord*.

In navolging van Bybee (2006, 2010) zouden we deze uitdrukkingen (idioms) en prefabs kunnen linken aan het cognitieve mechanisme *chunking*:

The underlying cognitive basis for morphosyntax and its hierarchical organization is the chunking of sequential experiences that occurs with repetition (Miller 1956; Newell 1990; Haiman 1994; Ellis 1996; Bybee 2002a). Chunking has been identified as a process influencing all cognitive systems, based on the general organization of the memory (Bybee 2010: 34).

Indien een werkwoordelijke uitdrukking, een samengesteld werkwoord en partikelwerkwoorden niet als meerdere woorden maar als één *chunk* worden opgeslagen, zijn ze makkelijker te produceren en te verwerken. De invloed van de werkwoorden op de woordvolgorde veronderstelt dan ook dat de grammatica en de grammaticale keuzes die de taalgebruiker maakt het resultaat zijn van het concrete taalgebruik, namelijk hoe vaak een taalgebruiker een bepaalde combinatie of *chunk* gebruikt en herhaalt waardoor deze *chunk* als een eenheid geactiveerd wordt. Dat betekent eveneens dat de grammatica niet los gezien kan worden van het taalgebruik: “grammar is the cognitive organization of one’s experience with language” (Bybee 2006: 711).

Hoewel woordfrequentie hier een sleutelbegrip is, bleek de frequentie van het werkwoord in deze studie geen effect te hebben. We hebben echter geen rekening gehouden met de verschillende betekenissen van het werkwoord. Het werkwoord *spelen* in de letterlijke betekenis kreeg dezelfde frequentie als het werkwoord *spelen* in de uitdrukking *ergens een rol in spelen*. In het vervolg zouden we een onderscheid moeten

⁷ Of lexicale bundels: “frequently recurring strings of words that often span traditional syntactic boundaries” (Tremblay et al. 2011)

maken tussen de frequentie van individuele werkwoorden, individuele werkwoorden met een vast opgeroepen voorzetsel of een partikel (prefabs), uitdrukkingen en samengestelde predicaten (idioms) en uitdrukkingen (of samengestelde predicaten) met een vast voorzetsel (prefabs). Daarnaast moeten we testen of reflexieve werkwoorden net zoals partikelwerkwoorden en samengestelde werkwoorden als een *chunk* opgeslagen zijn.

Tot slot kan ook het type VzC in combinatie met de werkwoorden die we onder de loep nemen een effect hebben op de woordvolgorde. De meerderheid van de VzC's bij een koppelwerkwoord zijn namelijk voorzetselvoorwerpen, in totaal 53% (bij een zelfstandig werkwoord daalt het aantal voorzetselvoorwerpen tot 43%). Een samengesteld werkwoord wordt zelfs in 82,5% van de gevallen met een voorzetselvoorwerp gecombineerd (voor de reflexieve, partikel- en combinatie werkwoorden is de combinatie met een voorzetselvoorwerp minder dan de helft, namelijk 43%). In paragraaf 5.2.1 hebben we vastgesteld dat voorzetselvoorwerpen net als de VzC's van *beperking* vaker in het achterveld staan. Hoewel we deze voorzetselvoorwerpen semantisch moeilijk kunnen onderverdelen – het voorzetsel wordt vast opgeroepen door het werkwoord waardoor het zijn letterlijke betekenis verliest (bijvoorbeeld het voorzetsel *op* in *gek zijn op iets*) – ligt de informatieve kern van het voorzetselvoorwerp niet op voorhand vast. Door deze voorzetselvoorwerpen in het achterveld te plaatsen, helpt de schrijver het verwerkingsproces van de toehoorder: het semantische zwaartepunt van de uitdrukking (het niet-werkwoordelijke deel van de uitdrukking of het gezegde) komt eerst, vervolgens het semantisch lege werkwoord en tot slot het informatieve en moeilijk voorspelbare voorzetselvoorwerp. Indien de schrijver het voorzetselvoorwerp voor het gezegde plaatst, moet de toehoorder eerst de informatie van het voorzetselvoorwerp verwerken en onmiddellijk erna het semantische zwaartepunt van de uitdrukking. Vanuit een lezergeoriënteerd standpunt lijkt de afwisseling tussen informatief prominente zinsdelen de beste en meest efficiënte volgorde en bij de keuze tussen twee volgordes wint, zoals al eerder bleek, de meest efficiënte (o.a. Hawkins 2004, Diessel 2011). Daarnaast sluit de volgorde *inherent element* < *werkwoord* < *voorzetselvoorwerp* ook aan bij Hawkins' verwerkingstheorie op basis van complexiteit en efficiëntie (Hawkins 1994, 2004). Zoals eerder vermeld, houdt deze theorie rekening met de afstand tussen gerelateerde woorden en hoeft het laatste woord nog niet bereikt te zijn om de hele zin te begrijpen. Bij de volgorde *voorzetselvoorwerp* < *inherent element* < *werkwoord* is de afstand van het voorzetselvoorwerp tot het werkwoord groter dan bij de volgorde met een voorzetselvoorwerp in het achterveld. De meest efficiënte volgorde bleek dan ook de volgorde te zijn waarbij het werkwoord niet uiterst rechts geplaatst wordt.

5.2.4 De discoursstatus

Voor de *discours- of informatiestatus van de VzC* hebben we rekening gehouden met de *bekendheid* en het *belang*, om op die manier de algemene tendens te testen dat oude informatie voor nieuwe staat (o.a. Gundel 1988; ANS 1997; Arnold et al. 2000) en onbelangrijke voor belangrijke (Grondelaers en Speelman 2007). We stelden vast dat deze twee operationalisaties interageerden. Als de referent in de VzC nieuwe informatie bevat, staat de VzC naar verwachting meer in achterveldpositie, maar deze trend is minder rechtdoorzee als de referent tevens minder belangrijk blijkt te zijn. Omgekeerd verkiest belangrijke informatie achterveldpositie, maar wordt dit effect afgeremd als de referent tevens bekend is (of bekender blijkt te zijn). De resultaten bevestigen dus de algemeen informationele tendensen (namelijk de *oud-voor-nieuw-regel* en *onbelangrijk-voor-belangrijk-regel*) en tonen ook aan hoe deze twee variabelen elkaar beïnvloeden.

Volgens ons is het interactie-effect het gevolg van coherentie en tekstverwerking (of zinsverwerking). Door een zin te beginnen met bekende, oude informatie kan de lezer een link leggen met wat er al eerder gezegd werd en ontstaat er coherentie⁸, wat de zin ook mentaal toegankelijker maakt: “The more grounding connections the clause has, the more mentally accessible it is, and thus the more coherent it is relative to the text in which it is embedded” (Givón 1995: 76). Tegelijkertijd vraagt onbelangrijke informatie aan het begin van de zin minder inspanningen van het werkgeheugen om deze informatie op te slaan, wat ook de mentale verwerking kan vergemakkelijken. Dat de twee operationalisaties elkaar beïnvloeden, kunnen we verklaren door coherentie in een tekst als een mentaal fenomeen te zien (Givón 1995). Tijdens het schrijfproces, de revisie en het editen, zal de schrijver zowel beïnvloed worden door zijn eigen mentale representatie, als door de mentale representatie van de lezer die de schrijver enkel kan veronderstellen. Coherentie ontstaat dus op basis van de mentale voorstelling die de schrijver heeft en door de verwerkingskost die de schrijver zelf ondervindt en veronderstelt.

De *oud-voor-nieuw-volgorde* en *onbelangrijk-voor-belangrijk-volgorde*, de manier waarop ze elkaar beïnvloeden en de verklaring in termen van coherentie en verwerking, stimuleren eveneens efficiënte communicatie (zie ook Arnold et al. 2000 voor gesproken communicatie). We weten ondertussen dat oude informatie veeleer vooraan in de zin staat, maar deze trend is minder duidelijk als de oude informatie belangrijk is. Nieuwe informatie staat voornamelijk achteraan in de zin, maar deze voorkeur neemt af naarmate de informatie onbelangrijk is. De lezer wordt bijgevolg geholpen in het leesproces. Indien nieuwe informatie toch in het middenveld staat, is de kans groot dat

⁸ Deze bevinding ligt ook in lijn met Jansen (1978). Jansen stelde vast dat onbepaalde VzC's meer in extrapositie staan dan bepaalde. Volgens Jansen is dat een “afspiegeling van de meer algemene tendens in natuurlijke talen om de zin te openen met elementen die wijzen naar dingen die aan spreker en hoorder bekend zijn, en de verwijzers naar onbekende zaken daar achter te zetten” (Jansen 1978: 82).

deze informatie niet belangrijk is. Tegelijkertijd kan oude informatie in het achterveld een indicatie zijn dat deze informatie wel belangrijk is.

Dit onderzoek toonde tot slot aan dat de operationalisatie in termen van *definietheid* geen rol speelt, in tegenstelling tot het onderzoek van Jansen (1978) en het onderzoek naar andere types van woordvolgordevariatie (o.a. Bresnan et al. 2007), en dat *complexiteit* niet ondergeschikt is aan de *discoursstatus* (in tegenstelling tot Givóns Code-Quantity Principle 1983 en Braecke 1990). Op basis van de multivariate analyse kunnen we besluiten dat *discoursstatus* en *complexiteit* een uniek en onafhankelijk effect hebben op de woordvolgordevariatie van VzC's in de journalistieke bijzin. Dat laatste ligt ook in lijn met wat Arnold et al. (2000) en Wasow (2002) vaststellen voor verschillende types van Engelse woordvolgordevariatie zoals *heavy NP-shift*, *dative alternation* en *verb-particle placement*.

5.3 Vervolgonderzoek

In deze studie hebben we vastgesteld dat meerdere factoren de variabele positie van VzC's in de Nederlandse bijzin bepalen. Het valt daarbij op dat deze positie niet enkel verklaard wordt door kenmerken van het variabele zinsdeel zelf, zoals *de semantische status van de VzC*, *de lengte van de VzC* of *de informatiestatus van de VzC*, maar dat ook andere variabelen, zoals *het werkwoord* en *de lengte van het middenveld* en *het achterveld*, een belangrijke rol spelen. Het idee dat verschillende variabelen aan de basis liggen, komt trouwens ook al voor in andere studies naar andere types van woordvolgordevariatie in het Nederlands en in andere talen, zoals in het Engels (o.a. Stallings, MacDonald & O'Seaghdha 1998; Wasow 1997).

Rekening houdend met het multifactoriële karakter van de woordvolgordevariatie, gaan we ervan uit dat verschillende mechanismen simultaan werkzaam zijn en dat er bijgevolg verschillende verklaringsprincipes simultaan geldig kunnen zijn. De functioneel-cognitieve verklaringen die we in de vorige paragraaf hebben voorgesteld, zijn dus niet de enig mogelijke verklaringen. Zeker de cognitieve interpretaties, zoals die in termen van een DLM-theorie en/of Hawkins' verwerkingstheorie, moeten in vervolgonderzoek verder geverifieerd worden. In vervolgonderzoek zullen de functioneel-cognitieve verklaringen dus psycho-experimenteel geverifieerd moeten worden. Functioneel-cognitieve verklaringen staan trouwens vaak ter discussie in de literatuur (o.a. Chomsky 2005) en krijgen anderzijds veel aandacht (o.a. Bybee 2007). We hopen met deze studie aangetoond te hebben dat er zeker ruimte is voor een dergelijke verklaring om woordvolgordevariatie te beschrijven, maar dat deze verklaring alleen ook onvoldoende is en dat er nog andere verklaringsmechanismen een rol kunnen spelen.

Vervolgonderzoek is altijd mogelijk en nodig, en hier is dat niet minder het geval. We denken aan drie verdere uitbreidingen. Een eerste mogelijkheid is om hetzelfde onderzoek op andere corpora uit te voeren. Verschillende corpora in eenzelfde taal kunnen namelijk significant van elkaar verschillen als het gaat over syntactische patronen (Biber 1993; Meyer 2002). Andere geschreven tekstgenres dan het journalistieke genre zouden dus een andere volgordevoorkeur kunnen hebben. Zuijdgeest (1995), bijvoorbeeld, gaf al aan dat juridische teksten vanwege hun vaak archaïsche taalgebruik en hun zware tangconstructies meer middenveldpositie zullen stimuleren.

Ten tweede zoomt deze studie naar woordvolgordevariatie in op één specifiek taalkundig woordvolgordefenomeen (namelijk de positie van VzC's in de Nederlandse journalistieke bijzin ingeleid door het grammaticaal onderschikkend voegwoord *dat*) en in veel mindere mate op de productieomstandigheden van de schrijver. Uitgaande van onze studie en op basis van de verschillende onderzochte variabelen kunnen we voorspellen welke volgorde taalgebruikers zullen kiezen. In een volgende stap zouden we deze voorspellingen kunnen vergelijken met de keuze die journalisten tijdens een

online-verwerking maken. Meer bepaald kunnen we nagaan of journalisten zinsdelen (ver)plaatsen op basis van de cognitieve-functionele verklaringen die we in de vorige paragrafen hebben besproken. Met deze stap wordt de plausibiliteit van het model nogmaals getest en kunnen we de innerlijke werking van taalverwerking en -productie van de journalist, en van de schrijver in het algemeen, blootleggen. Experimenteel onderzoek (met bijvoorbeeld eye-tracking technieken) is bijgevolg nodig om een beter zicht te krijgen op dit schrijfproces.

Tot slot kan deze studie ook de aanzet geven tot een benadering vanuit een probabilistisch netwerk (bv. een Bayesiaans model (Pearl 1988)) en memory-based learning (Daelemans & Van den Bosch 2005). Logistische-regressieanalyse werd in het verleden meermaals gebruikt omdat het tot interessante inzichten heeft geleid die de syntactische keuze van taalgebruikers kan motiveren (zie o.a. Harrell 2001; Baayen 2008; Speelman 2014 voor een introductie en zie ook Theijssen et al. 2013 voor een opsomming van de voordelen van verschillende statistische modellen zoals multivariate regressieanalyse, een Bayesiaans model en memory-based learning). De meeste studies, waaronder ook deze studie, maken daarbij gebruik van zogenaamde *higher-level* variabelen, die syntactisch van aard kunnen zijn (bv. *grammaticale status van de VzC*), semantisch (bv. *semantische status van de VzC*) of discoursgerelateerd (bv. *toegankelijkheid van de VzC*). Een nadeel is dat deze kenmerken meestal manueel geannoteerd moeten worden. De tussenkomst van de onderzoeker zal onvermijdelijk zorgen voor enige subjectiviteit bij de operationalisatie van deze variabelen. Een groter nadeel volgens o.a. Theijssen et al. (2013) is dat een regressiemodel niet in staat is om cognitieve aspecten te integreren.

Multivariate models such as regression models can successfully be exploited for the purpose of analysing the relative importance of higher-level features in specific data sets under investigation, but these models cannot elucidate the role of these features in cognitive processes. (Theijssen et al. 2013: 230)

Een mogelijke oplossing is een probabilistisch netwerk, zoals een Bayesiaans model. Zo'n netwerk kan opgebouwd worden op basis van de *higher-level* kenmerken die we in onderhavig onderzoek geannoteerd en geanalyseerd hebben. In tegenstelling tot een logistisch regressiemodel is het echter niet gebaseerd op de statistische verdeling van deze predictoren en kan het ook niet-significante predictoren en zelfs ruis meenemen in het model:

Integrating partial and noisy sensory input and volatile procedural and semantic memory is what the brain does all the time, especially in the initial stages of the processing where not all information is available yet. Therefore, Bayesian Networks form an attractive analogue for cognitive processes. (Theijssen et al. 2013)

Memory-based learning gaat nog een stapje verder en gebruikt enkel lexicale informatie. Het is een computationele aanpak om natuurlijke taalverwerkingsproblemen op te lossen. De aanpak is gebaseerd op een combinatie van een geheugenonderdeel en een taalverwerkingsonderdeel. Een taal leer je namelijk door woorden, woordconstructies en eventuele constructionele problemen op te slaan in het geheugen. Deze aanpak is dus een vertaling van de taalgebruiksgebaseerde aanpak en veronderstelt geen kennis van *higher-level* kenmerken. In deze aanpak zou het werkwoord een cruciale rol kunnen spelen. Theijssen et al. (2013), bijvoorbeeld, toont aan dat een memory-based model voor de datiefalternantie een accuraat model is, vergelijkbaar met een logistisch regressiemodel en een Bayesiaans model, en stelt verder vast dat verschillende werkwoorden hun eigen woordvolgordevoorkeur hebben, “which might give credibility to a theory based on memory-based processes” (Theijssen et al. 2013: 258).

Bibliografie

- Agresti, A. 2007. *An Introduction to Categorical Data Analysis. Second edition*. New York: John Wiley and Sons.
- Allum, P.H. & Wheeldon, L.R. 2007. 'Planning Scope in Spoken Sentence Production: The Role of Grammatical Units'. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33 (4), 791-810.
- Altmann, G.T.M. & Kamide, Y. 1999. 'Incremental interpretation at verbs: Restricting the domain of subsequent reference'. *Cognition*, 73 (3), 247-264.
- ANS – *Algemene Nederlandse Spraakkunst*. 1997. Haeseryn, W., Romijn, K., Geerts, G., de Rooij, J. & van den Toorn, M.C. (red.). Groningen/Deurne: Martinus Nijhoff – Wolters Plantyn.
- Ariel, M. 1990. *Accessing noun phrase antecedents*. London & New York: Routledge.
- Arnold, J.E., Wasow, T., Losongco, A. & Ginstrom, R. 2000. 'Heaviness vs. newness: the effects of structural complexity and discourse status on constituent ordering'. *Language*, 76, 28-55.
- Baayen, R. H., Feldman, L. B. & Schreuder, R. 2006. 'Morphological influences on the recognition of monosyllabic monomorphemic words'. *Journal of Memory and Language*, 53, 496-512.
- Baayen, R. H. 2008. *Analyzing Linguistic Data. A Practical Introduction to Statistics Using R*. Cambridge: Cambridge University Press.
- Balota, D. A., Cortese, M. J., Sargent-Marshall, S., Spieler, D. H. & Yap, M. J. 2004. 'Visual word recognition of single-syllable words'. *Journal of Experimental Psychology: General*, 133 (2), 283-316.
- Baltin, M. 1981. 'Strict bounding'. In: C. Baker and J. McCarthy (eds.). *The logical problem of language acquisition*. Cambridge/London: MIT Press.
- Baltin, M. 1983. 'Extraposition: Bounding versus government-binding'. *Linguistic Inquiry*, 14, 155-162.
- Baltin, M. 1984. 'Extraposition rules and discontinuous constituents'. *Linguistic Inquiry*, 15, 157-163.
- Barbiers, S. 1995. *The Syntax of Interpretation*. Doctoral Dissertation, University of Leyden, Leiden.
- Barbiers, S. 2000. 'The right periphery in SOV languages: English and Dutch'. In: Peter Svenonius, 181-218.
- Barbiers, S. 2008. 'Werkwoordclusters en de grammatica van de rechterperiferie'. *Nederlandse Taalkunde*, 13 (2), 160-187.
- Bartsch, R. & Vennemann, Th. 1972. *Semantic Structures*. Frankfurt/M.
- Beaman, K. 1984. 'Coordination and Subordination Revisited: Syntactic Complexity in Spoken and Written Narrative Discourse'. In: D. Tannen and R. Freedle (Eds.). *Coherence in Spoken and Written Discourse*. Norwood, NJ: Ablex, 45-80.
- Behaghel, O. 1909. 'Beziehungen zwischen Umfang und Reihenfolge von Satzgliedern'. *Indogermanische Forschungen*, 25, 110-142.

- Behaghel, O. 1930. 'Von deutscher Wortstellung'. *Zeitschrift für Deutschkunde*, 44, 81-89.
- Bever, T. 1970. 'The Cognitive Basis for Linguistic Structures'. In: J.R. Hayes (ed.). *Cognition and the Development of Language*. New York.
- Biber, D. 1993. 'Representativeness in corpus design'. *Literary and Linguistic Computing*, 8 (4), 243-257.
- Booij, G.E. 1973-74. 'Zinsbepalingen in het Nederlands.' *Spektator*, 3, 619-646.
- Bouman, J. 2009. *Vlot en helder schrijven*. Alphen aan den Rijn: Kluwer.
- Braecke, C. 1982-1983. De positie van voorzetselgroepen in tangconstructies: een bijdrage tot de empirische studie van sprekersstrategieën. Proefschrift RUG (Rijksuniversiteit Gent).
- Braecke, C. 1990. "Uit de tang" of [+ or -] prominent?. *Taal en Tongval*, 3, 125-134.
- Bresnan, J., Cueni, A., Nikitina, T. & Baayen, R.H. 2005. 'Predicting the Dative Alternation'. KNAW Academy Colloquium: Cognitive Foundations of Interpretation. October 27-28, 2004. Amsterdam.
- Bresnan, J. W. & Ford, M. 2010. 'Predicting syntax: Processing dative constructions in American and Australian varieties of English'. *Language*, 86 (1), 168-213.
- Broekhuis, H. 2001. 'Hoeveel syntaxis zit er in het lexicon'. *Nederlandse Taalkunde*, 6, 3-21.
- Broekhuis, H. 2005. 'Generatieve Syntaxis. De balans tussen eenvoud en complexiteit'. *Nederlandse taalkunde*, 10 (3/4), 223-245.
- Broekhuis, H. 2008. *Derivations and Evaluations: Object Shift in the Germanic Languages*. Berlin/New York: Mouton de Gruyter.
- Broekhuis, H., Corver, N., Vos, R. & Bennis, H. 2016. *Syntax of Dutch: verbs and verb phrases*. Amsterdam: Amsterdam University Press.
- Burridge, K. 1993. *Syntactic change in Germanic. Aspects of language change in Germanic with particular reference to Middle Dutch*. Amsterdam: John Benjamins.
- Bybee, J. 1985. *Morphology: A study into the relation between meaning and form*. Amsterdam: John Benjamins.
- Bybee, J. 2006. 'From usage to grammar: the mind's response to repetition'. *Language*, 82, 711-733.
- Bybee, J. 2007. *Frequency of use and the organization of language*. Oxford: Oxford University Press.
- Bybee, J. 2010. *Language, usage and cognition*. Cambridge: Cambridge University Press.
- Bybee, J. 2013. 'Usage-based Theory and exemplar presentations of constructions'. In: T. Hoffman and G. Trousdale (eds.). *The Oxford Handbook of Construction Grammar*. Oxford: Oxford University Press, 49-69.
- Bybee, J. & Hopper, P. 2001. 'Introduction to frequency and the emergence of linguistic structure'. In: J. Bybee and P. Hopper (eds.). *Frequency and the emergence of linguistic structure*. Amsterdam: John Benjamins, 1-24.
- Bybee, J. & Beckner, C. 2011. 'Usage-based Theory'. In: B. Heine and H. Narrog (eds.). *The Oxford Handbook of Linguistic Analysis*. Oxford: Oxford University Press, 826-855.
- Chomsky, N. 1992. 'A Minimalist Program for linguistic theory'. *MIT Occasional Papers in Linguistics*, 1.
- Chomsky, N. 1995. *The minimalist program*. Cambridge: MIT Press.
- Chomsky, N. 2005. 'Three factors in language design'. *Linguistic Inquiry*, 36 (1), 1-22.
- Colleman, T. 2010. 'Beyond the dative alternation: the semantics of the Dutch aan-dative'. In: D. Glynn and K. Fischer (Eds.). *Quantitative Methods in Cognitive Semantics: Corpus-driven Approaches*. Berlin/New York: Mouton de Gruyter, 271-303.
- Combinatiewoordenboek: van Nederlandse substantieven (zelfstandige naamwoorden) met hun vaste verba (werkwoorden)* – 2006. P. De Kleijn. Amsterdam: Rozenberg Publishers.
- Coussé, E. 2003. 'Volgordevariatie en herinterpretatie van de tweeledige werkwoordelijke eindgroep in de bijzin'. *Taal en Tongval*, 55, 138-156.
- Coussé, E. 2008. Motivaties voor volgordevariatie. Een diachrone studie van werkwoordsvolgorde in het Nederlands. Proefschrift, Universiteit Gent.
- Coussé, E. 2009. 'Focus, complexiteit en extrapositie. Over de veranderende woordvolgorde in het Nederlands'. *Neerlandistiek.nl*, 9 (4), 1-32.

- Coussé, E. 2011. 'De extrapositie van het subject in passieve zinnen. Diachrone trends en motivaties'. *Nederlandse Taalkunde*, 16 (special issue The right periphery of the Dutch sentence), 245-263.
- Coussé, E., Arfs, M. & De Sutter, G. 2008. 'Variabele werkwoordsvolgorde in de Nederlandse werkwoordelijke eindgroep. Een taalgebruiksgebaseerd perspectief op de synchronie en diachronie van de zgn. rode en groene woordvolgorde'. In: G. Rawoens (Ed.). *Taal aan den lijve. Het gebruik van corpora in taalkundig onderzoek en taalonderwijs*. Ghent: Academia Press, 29-47.
- Croft, W. 2001. *Radical construction grammar. Syntactic theory in typological perspective*. Oxford: Oxford university press.
- Culicover, P. & Rochemont, M. 1990. 'Extrapolation and the complement principle'. *Linguistic Inquiry*, 21, 23-47.
- Daelemans, W. & van den Bosch, A. 2005. *Memory-Based Language Processing*. Cambridge: Cambridge University Press.
- De Haan, Ger. 1974. 'On extraposition'. *Spektator*, 4, 161-183.
- De Haan, Ger. 1976. 'Regelordening en domeinformulering op transformaties'. In: Evers en Koefoed (eds.), 279-302.
- De Haan, Ger. 1979. *Condition on rules: the proper balance between syntax and semantics*. Dordrecht: Foris.
- Delaere, I. 2015. Do translations walk the line? Visually exploring translated and non-translated texts in search of norm conformity. Proefschrift, Universiteit Gent.
- De Meersman, A. 1980. 'Woordvolgorde in 14de-eeuws Brabants'. Gent: Koninklijke academie voor Nederlandse taal- en letterkunde.
- De Meersman, A. 1988. 'In en uit de tang in de veertiende eeuw: de plaats van het direct object in de bijzin: corpusonderzoek tegen een structurele, taaltypologische en pragmatische achtergrond'. In: J.A. van Leuvensteijn (red.). *Uitgangspunten en toepassingen: taalkundige studies over Middelnederlands en zestiende- en zeventiende eeuws Nederlands*. Amsterdam, 95-124.
- Den Hertog, C.H. 1973. *Nederlandse spraakkunst*. Amsterdam: Versluys.
- De Schutter, G. 1976. 'De bouw van de Nederlandse zin. Beschrijving en voorstel tot beregeling'. Verslagen en mededelingen van de Koninklijke Academie voor Taal en Letterkunde. Gent: KANTL.
- De Schutter, G. 1985. 'Pragmatic and syntactic aspects of word order in Dutch'. In: A.M. Bolkestein e.a. (reds.). *Syntax and pragmatics in Functional Grammar*. Dordrecht: Foris, 137-154.
- De Schutter, G. 1988. 'In en uit de tang in de Middelnederlandse bijzin'. *Leuvense Bijdragen*, 77, 385-401.
- De Schutter, G. 1995. 'Extrapositie in het Middelnederlands. Proza- en rijmteksten als informatiebron voor syntactisch onderzoek'. *Taal en Tongval*, 47, 195-204.
- De Schutter, G. 2003. 'De zinsbouw in het vroege Middelnederlands. Een case study'. *Taal en Tongval*, 55, 41-67.
- De Schutter, G. & Van Hauwermeiren, P. 1983. *De structuur van het Nederlands: taalbeschouwelijke grammatica*. Malle: De Sikkell.
- De Sutter, G. 2007. 'Naar een corpusgebaseerde, cognitief-functionele verklaring van de woordvolgordevariatie in tweeledige werkwoordelijke eindgroepen'. *Nederlandse Taalkunde*, 12, 302-330.
- De Sutter, G., Spielman, D. & Geeraerts, D. 2007. 'Luisteren schrijvers naar hun innerlijke stem? De invloed van ritmische factoren op de woordvolgorde in geschreven werkwoordelijke eindgroepen'. *Neerlandistiek.nl*, artikelnummer 07.01.
- De Sutter, G. & Van de Velde, M. 2008. 'Do the mechanisms that govern syntactic choices differ between original and translated language? A corpus-based translation study of PP extraposition in Dutch and German'. *Proceedings of the international symposium on using corpora in contrastive and translation studies (UCCTS 2008)*.
- De Vries, M. 2009. 'The left and right periphery in Dutch'. *Linguistic review*, 26 (2/3), 291-327.

- De Vries, M. 2011. 'Extrapositie'. *Nederlandse Taalkunde*, 16, 273 - 295.
- Diessel, H. 2011. 'Review article of 'Language, usage and cognition' by Joan Bybee'. *Language* 87, 830-844.
- Dik, S.C. 1978. *Functional Grammar*. Amsterdam: North-Holland.
- Ellis, N. C. 2002a. 'Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition'. *Studies in Second Language Acquisition*, 24 (2), 143-188.
- Emonds, J. 1976. *A Transformational Approach to English Syntax. Root, Structure-Preserving and Local Transformations*. New York: Academic Press.
- Emonds, J. 1985. *A unified theory of syntactic categories*. Dordrecht: Foris.
- Erman, B. & Warren, B. 2000. 'The idiom principle and the open choice principle'. *Text*, 20 (1), 29-62.
- Ferreira, F. & Swets, B. 2002. 'How Incremental Is Language Production? Evidence from the Production of Utterances Requiring the Computation of Arithmetic Sums'. *Journal of Memory and Language*, 46, 57-84.
- Fillmore, C. J. 1968. 'The Case for Case'. In: E. Bach and R. Harms (Ed.). *Universals in Linguistic Theory*. New York: Holt, Rinehart and Winston, 1-88.
- Flower, L. & Hayes, J.R. 1981. 'A Cognitive Process Theory of Writing'. *College Composition and Communication*, 32, 365-387.
- Ford, M. & Holmes, V.M. 1978. 'Planning units in sentence production'. *Cognition*, 6, 35-53.
- Fox, J. 2003. 'Effect displays in R for generalised linear models'. *Journal of Statistical Software*, 8 (15), 1-27.
- Fox, J. & Hong, J. 2009. 'Effect displays in R for multinomial and proportional-odds logit models. Extensions to the effects package'. *Journal of Statistical Software*, 32 (1), 1-24.
- Gibson, E. 1998. 'Linguistic complexity: locality of syntactic dependencies'. *Cognition*, 68, 1-76.
- Gibson, E. 2000. 'The dependency locality theory: A distance-based theory of linguistic complexity'. In: Y. Miyashita, A. Marantz, and W. O'Neil (Eds.). *Image, Language, Brain*. Cambridge, MA: MIT Press, 95-126.
- Gildea, D. & Temperley, D. 2010. 'Do Grammars Minimize Dependency Length?'. *Cognitive Science* 34, 286-310.
- Givón, T. 1979. *On understanding grammar*. New York: Academic Press.
- Givón, T. 1983. *Topic Continuity in Discourse: A Quantitative Cross Language Study*. Amsterdam: John Benjamins.
- Givón, T. 1988. 'The pragmatics of word-order: predictability, importance and attention'. In: M. Hammond e.a. (eds.). *Studies in syntactic typology*. Amsterdam: John Benjamins, 243-284.
- Givón, T. 1991. 'Isomorphism in the grammatical code: cognitive and biological considerations'. *Studies in Language*, 15, 85-114.
- Givón, T. 1995. *Functionalism and grammar*. Philadelphia: John Benjamins.
- Gries, S.Th. 2002a. 'The influence of processing on grammatical variation: particle placement in English'. In: Nicole Dehé, Ray Jackendoff, Andrew McIntyre, & Silke Urban (eds.). *Verb-particle explorations*. Berlin & New York: Mouton de Gruyter, 269-288.
- Gries, S.Th. 2004. *Multifactorial analysis in corpus linguistics: a study of Particle Placement*. Peking: Peking University Press.
- Gries, S.Th. 2013. *Statistics for linguistics with R. 2nd rev. and ext. ed.* Berlin & New York: De Gruyter Mouton.
- Gries, S.Th. & Stefanowitsch, A. 2004. 'Extending collocation analysis: a corpus-based perspective on 'alternations''. *International Journal of Corpus Linguistics*, 9 (1), 97-129.
- Grondelaers, S. 2000. *De distributie van niet-anaforisch er buiten de eerste zinsplaats. Sociolexicologische, functionele en psycholinguïstische aspecten van er's status als presentatief signaal*. Unpublished doctoral dissertation, University of Leuven.

- Grondelaers, S. & Speelman, D. 2007. 'A variationist account of constituent ordering in presentative sentences in Belgian Dutch'. *Corpus Linguistics and Linguistic Theory*, 3, 161-193.
- Grondelaers, S., Speelman, D. & Geeraerts, D. 2008. 'National variation in the use of er "there". Regional and diachronic constraints on cognitive explanations'. In: Gitte Kristiansen and René Dirven (ed.). *Cognitive Sociolinguistics. Language Variation, Cultural Models, Social Systems*. Berlin/New York: Mouton de Gruyter, 153-203.
- Grondelaers, S., Speelman, D., Drieghe, D., Brysbaert, M. & Geeraerts, D. 2009. 'Introducing a new entity into discourse: Comprehension and production evidence for the status of Dutch er "there" as a higher-level expectancy monitor'. *Acta Psychologica*, 130, 153-160.
- Guéron, J. 1976. 'Interpretation of PP complements: A semantic filter for PP extraposition'. In H.C. van Riemsdijk (ed.). *Green ideas blown up*. Publikaties van het instituut voor Algemene Taalwetenschap UvA, 13.
- Guéron, J. 1980. 'On the syntax and semantics of PP extraposition'. *Linguistic Inquiry*, 11 (4), 637-678.
- Gundel, J. K. 1988. 'Universals of topic-comment structure'. In: M. Hammond, E. Moravcsik and J. Wirth (eds.). *Studies in Syntactic Typology*. Amsterdam: John Benjamins, 209-239.
- Haeseryn, W. 1997. 'Achteropplaatsing van elementen in de zin'. *Colloquium Neerlandicum*, 13, 303-326.
- Haeseryn, W. 1998. 'Achteropplaatsing van elementen in de zin'. In: Brems, Hugo e.a. (red.). *Nederlands 200 jaar later. Handelingen Dertiende Colloquium Neerlandicum* (Leiden 1997). Woubrugge: IVN, 303-326.
- Haider, H. 1997. 'Precedence among predicates'. *Journal of Comparative Germanic Linguistics*, 1, 3-41.
- Haider, H. 2000. 'OV is more basic than VO'. In: Peter Svenonius, 45-67.
- Harrell, F.E. 2001. *Regression modeling strategies: with applications to linear models, logistic regression and survival analysis*. New York: Springer.
- Hartsuiker, R.J. & Westenberg, C. 2000. 'Word order priming in written and spoken sentence production'. *Cognition*, 75, B27-B39.
- Hawkins, J.A. 1990. 'A parsing theory of word order universals'. *Linguistic Inquiry*, 21, 223-261.
- Hawkins, J.A. 1994. *A performance theory of order and constituency*. Cambridge: Cambridge University Press.
- Hawkins, J.A. 2004. *Efficiency and Complexity in Grammars*. Oxford: Oxford university press.
- Hengeveld, K. & Mackenzie, J.L. 2008. *Functional Discourse Grammar. A Typologically-Based Theory of Language Structure*. Oxford/New York: Oxford University Press.
- Hoeksema, J. 2008. 'Van + bijwoord van polariteit: een geval van verplichte extrapositie?'. *Tabu*, 37 (1/2), 69-78.
- Hoeksema, J. 2014. 'De plaats van het voorzetselvoorwerp'. *Nederlandse Taalkunde*, 19, 221-244.
- Hoekstra, T. 1984. *Transitivity: grammatical relations in government-binding theory*. Dordrecht: Foris.
- Hoekstra, T. 1987. 'Extraposition and SOV'. *Tabu*, 17, 133-142.
- Holmes, V.M., Stowe, L.A. & Cupples, L. 1989. 'Lexical expectations in parsing complement verb sentences'. *Journal of Memory & Language*, 28, 668-689.
- Jansen, F. 1978. 'Hoe krijgt een spreker zijn woorden op een rijtje? Taalgebruikaspecten van de 'PP over V' konstruktie'. In: J. G. Kooij (ed.). *Aspekten van de woordvolgorde in het Nederlands*. Leiden: Vakgroep Nederlandse Taal- and Letterkunde, 70-104.
- Jansen, F. 1979. 'On tracing conditioning factors of movement rules: extraposition of PP in spoken Dutch'. In: M. Van de Velde and W. Vandeweghe (ed.). *Sprachstruktur, Individuum und Gesellschaft*. Tübingen: Niemeyer, 83-93.
- Jansen, F. 2003. 'Sentence processing theories and the position of complex constituents in Dutch texts'. In: L. Cornips (eds.). *Linguistics in the Netherlands 2003*. Amsterdam/Philadelphia: John Benjamins, 83-92.

- Jansen, F. & Wijnands, R. 2004. 'Doorkruisingen van het Links-Rechtsprincipe'. Neerlandistiek.nl, artikelnummer 04.01.
- Kaan, E. 1993. 'Extraposition from NP in Dutch: Some consequences of Minimalism. GAGL, 36, 144-151.
- Kamide, Y., Altmann, G.T.M. & Haywood, S.L. 2003. 'The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements'. *Journal of Memory and Language*, 49, 133-159.
- Karreman, J., Enschoot, R. & Schuur, U. 2013. *Tekstanalyse: methoden en toepassingen*. Assen: Van Gorcum.
- Kayne, R. 1992. *Word order*. Lisbon: Glow lecture.
- Kayne, R. 1994. *The Antisymmetry of Syntax*. Cambridge, Mass.: MIT Press.
- Kerstens, J.G. & Sturm, A.N. 1979. 'Over problemen met intuities'. *De nieuwe taalgids*, 72 (5), 449-465.
- Keuleers, E., Brysbaert, M. & New, B. 2010. 'SUBTLEX-NL: A new measure for Dutch word frequency based on film subtitles'. *Behavior Research Methods*, 42 (3), 643-650.
- Kidd, E., Lieven, E. & Tomasello, M. 2006. 'Examining the Role of Lexical Frequency in Children's Processing and Acquisition of sentential Complements'. *Cognitive Development*, 21, 93-107.
- Kooij, J. 1973. *Is Nederlands een SOV Taal?* Amsterdam. (Oratie Leiden).
- Koops, A. 1984-85. 'De zogenaamde PP-over-V constructie over Aaldrik Koops'. *Spektator*, 14, 352-357.
- Koster, J. 1973. 'PP over V en de theorie van J. Emonds'. *Spektator*, 2, 294-309.
- Koster, J. 1974. 'Het werkwoord als spiegelcentrum'. *Spektator*, 3, 601-618.
- Koster, J. 1975. 'Dutch as a SOV Language'. *Linguistic Analysis*, 1, 111-136.
- Koster, J. 1978. *Locality principles in syntax*. Dordrecht: Foris.
- Koster, J. 1994. 'Predicate Incorporation and the Word Order of Dutch'. In: ed. G. Cinque, J. Koster, J.Y. Pollock, L. Rizzi and R. Zanuttini. *Paths Towards Universal Grammar: Studies in Honor of Richard S. Kayne*. Washington DC: Georgetown University Press.
- Koster, J. 1999b. 'Empty Objects in Dutch'. Ms., University of Groningen.
- Koster, J. 1999c. 'The Word Orders of English and Dutch: Collective vs. Individual Checking'. In: W. Abraham (ed.). *Groninger Arbeiten zur germanistischen Linguistik*. Groningen: University of Groningen, 1-42.
- Koster, J. 2000a. 'Pied Piping and The Word Orders of English and Dutch'. In: M. Hirotani, A. Coetzee, N. Hall and J.-Y. Kim (eds.). *NELS 30: Proceedings of the North East Linguistic Society*. Amherst, MA: GLSA, 415-426.
- Koster, J. 2000b. 'Extraposition as Parallel Construal'. Ms., University of Groningen.
- Koster, J. 2001b. 'Mirror Symmetry in Dutch'. In: Marc van Oostendorp and Elena Anagnostopoulou (eds.). *Grammar in Progress: Articles Presented at the 20th Anniversary of the Comparison of Grammatical Models Group in Tilburg*. Roquade, Amsterdam.
- Labov, W. 1972. *Sociolinguistic patterns*. Oxford: Blackwell.
- Lavid, J., Arús, J. & Moratón, L. 2012. 'Genre Realized in Theme: The case of news reports and commentaries. *Discours: Revue de linguistique, psycholinguistique et informatique*. Retrieved from <http://discours.revues.org/8623>.
- Levelt, W.J.M. 1989. *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- Levin, B. 1995. 'Approaches to Lexical Semantic Representation'. In: D. Walker, A. Zampolli and N. Calzolari (eds.). *Automating the Lexicon I: Research and Practice in a Multilingual Environment*. Oxford: Oxford University Press, 53-91.
- MacDonald, M. 1994. 'Probabilistic constraints and syntactic ambiguity resolution'. *Language & Cognitive Processes*, 9, 157-201.
- Macken, L., Declercq, O. & Paulussen, H. 2011. 'Dutch Parallel Corpus: a Balanced Copyright-Cleared Parallel Corpus'. *META*, 56 (2), 374-390.

- Maks, I., van der Vliet, H. Görög, A. & Vossen, P. 2013. User Documentation of Cornetto LMF. Lexical Resource for Dutch. Amsterdam: Vrije Universiteit, CLARIN project.
- Martin, W. & Maks, J. 2005. Referentie Bestand Nederlands. Documentatie. Amsterdam: Vrije Universiteit, Research Group Lexicology/Terminology.
- Meyer, C.F. 2002. *English corpus linguistics: an introduction*. Cambridge: Cambridge University Press.
- Nelder, J.A. & Wedderburn, R.W.M. 1972. 'Generalized Linear Models'. *Journal of the Royal Statistical Society. Series A (General)*, 135 (3), 370-384.
- Newmeyer, F. 2003a. 'Grammar is grammar and usage is usage'. *Language*, 79, 682-707.
- Olohan, M. & Baker, M. 2000. 'Reporting that in Translated English: Evidence for Subliminal Processes of Explicitation?'. *Across Languages and Cultures*, 1 (2), 141-158.
- Pearl, J. 1988. *Probabilistic reasoning in intelligent systems: Networks of Plausible Inference*. Cambridge, MA: Morgan Kaufmann Publishers Inc.
- Perrin, D. & Ehrensberger-Dow, M. 2006. 'Journalists' Language Awareness: Inferences from Writing Strategies'. *Revista alicantina de Estudios Ingleses*, 19, 319-343.
- Quirk, R., Greenbaum, S., Leech, G. & Svartvik, J. 1985. *A Comprehensive Grammar of the English Language*. London: Longman.
- R Development Core Team – 2008. R: A language and environment for statistical computing. Vienna, Austria. Available from <http://www.R-project.org>.
- Renkema, J. 1989. 'Vermijd tangconstructies'. *Communicatief*, 2, 34-40.
- Renkema, J. 1989. 'Tangconstructies, experimenteel onderzoek naar leesbaarheid en attentiewaarde'. *Spektator*, 12, 444-462.
- Rietveld, T. & van Hout, R. 2005. *Statistics in Language Research: Analysis of Variance*. Berlin: Mouton de Gruyter.
- Rizzi, L. 1989. 'Relativized Minimality'. Ms. University of Geneva.
- Rochmont, M. & Culicover, P. 1990. *English focus constructions and the theory of grammar*. Cambridge: Cambridge University Press.
- Ross, J.R. 1973. 'The Penthouse Principle and the Order of Constituents'. In: You take the High Node and I'll take the Low Node. Chicago: CLS, 397-422.
- Shih, S. & Grafmiller, J. 2010. 'Weighing in on end weight'. Presentation at the workshop Development of Syntactic Alternations. Stanford University, 12 Nov.
- Shih, S. & Grafmiller, J. 2011. 'Weighing in on end weight'. Annual Meeting of the Linguistic Society of America.
- Speelman, D. 2014. 'Logistic regression: A confirmatory technique for comparisons in corpus linguistics'. In: D. Glynn and J.A. Robinson (ed.). *Corpus Methods for Semantics: Quantitative studies in polysemy and synonymy*. Amsterdam: John Benjamins, 487-533.
- Stahl S.A. & Nagy W.E. 2006. *Teaching Word Meanings*. Mahwah, New Jersey: Lawrence Erlbaum.
- Stallings, L.M., O'seaghdha, P.G & MacDonald, M.C. 1998. 'Phrasal ordering constraints in sentence production: Phrase length and verb disposition in heavy NP shift'. *Journal of Memory and Language*, 39 (3), 392-417.
- Stowell, T. 1981. Origins of phrase structure. Thesis (Ph.D.), Massachusetts Institute of Technology.
- Szmrecsanyi, B. 2004. 'On operationalizing syntactic complexity'. In: G. Purnelle, C. Fairon and A. Dister (eds.). *Le poids des mots. Proceedings of the 7th International Conference on Textual Data Statistical Analysis*. March 10-12, 2004, Vol. 2. Louvain-la-Neuve: Presses universitaires de Louvain, 1032-1039.
- Szmrecsanyi, B. 2006. *Morphosyntactic persistence in spoken English. A corpus study at the intersection of variationist sociolinguistics, psycholinguistics, and discourse analysis*. Berlin/New York: Mouton de Gruyter.
- Temperley, D. 2007. 'Minimization of dependency length in written English'. *Cognition*, 105 (2), 300-333.

- Theijssen, D., ten Bosch, L., Boves, L., Cranen, B. & van Halteren, H. 2013. 'Choosing alternatives: Using Bayesian Networks and memory-based learning to study the dative alternation'. *Corpus Linguistics and Linguistic Theory*, 9 (2), 227-262.
- Tily, H., Hemforth, B., Arnon, I., Shuval, N., Snider, N. & Wasow, T. 2008. 'Eye movements reflect comprehenders' knowledge of syntactic structure probability'. Presentation. *Architectures and Mechanisms for Language Processing*. UK: Cambridge.
- Tremblay, A., Derwing, B., Libben, G. & Westbury, C. 2011. 'Processing advantages of lexical bundles: Evidence from self-paced reading and sentence recall tasks'. *Language Learning*, 61 (2), 569-613.
- Uhlenbeck, E.M. 1962. 'De beginselen van het syntactisch onderzoek'. *Taalonderzoek in onze tijd*. Den Haag: Servire, 17-39.
- Vanacker, V.F. 1973. 'Een onnederlandse konstruktie in zuidwestelijke dialecten'. *Album Willem Pée*. Tongeren, 367-377.
- Vanacker, V.F. 1977. 'Syntactische overeenkomsten tussen Frans-Vlaamse en Westvlaamse dialecten'. *De Franse Nederlanden. Jaarboek uitgegeven door de Stichting Ons Erfdeel*, 2, 206-215.
- Van Bergen, G. & P. de Swart. 2010. 'Scrambling in Spoken Dutch. Definiteness versus weight as determinants of word order variation'. *Corpus Linguistics and Linguistic Theory*, 6 (2), 267-295.
- Van Canegem-Ardijns, I. 2004. The use of prepositional objects of adjectives in brace- and extraposition constructions in Dutch. *Proefschrift*, K.U. Leuven.
- Van Canegem-Ardijns, I. 2006. 'The extraposition of prepositional objects of adjectives in Dutch'. *Linguistics*, 44, 425-457.
- Van Dale groot woordenboek van hedendaags Nederlands – 1991. P.G.J. van Sterkenburg & G.E. Booij (red.). Utrecht: Van Dale Lexicografie.
- Van den Berg, B. 1949. 'De zinsbouw in het Nederlands'. *De Nieuwe Taalgids*, 42, 120-124 en 205-209.
- Van den Berg, B. 1952. *Beknopte Nederlandse Spraakkunst voor voortgezet onderwijs en particulier gebruik*. 's-Gravenhage: Van Goor.
- Van den Berg, E. 1991. 'Over de syntaxis van Middelnederlandse rijmtteksten Evert van den Berg'. *Tabu*, 21 (2), 67-75.
- Vandenbosch, L. 1985. 'PPoverV: een functioneel-pragmatisch principe'. *Antwerp Papers in Linguistics*, 46, 1-72.
- Vanden Wyngaerd, G. 1989. 'Object Shift as an A-movement Rule'. *MIT Working Papers in Linguistics* 11, 256-271.
- Van der Horst, J.M. 2008. *Geschiedenis van de Nederlandse syntaxis*. Leuven: Universitaire Pers Leuven.
- Van de Velde, F. 2009. 'PP extraction and extraposition in FDG'. *Second FDG workshop: The morphosyntactic level*. Oviedo.
- Van de Velde, F. 2012. 'PP extraction and extraposition in Functional Discourse Grammar'. *Language Sciences*, 34 (4), 433-454.
- Van de Velde, M. 1973. 'De Ausklammerung in het Duits en in het Nederlands'. *Studia Germanica Gandensia*, 14, 119-142.
- Vandevoorde, L. 2016. On semantic differences: a multivariate corpus-based study of the semantic field of inchoativity in translated and non-translated Dutch. *Proefschrift*, Universiteit Gent.
- Vandeweghe, W. & Devos, M. 2000. *Grammatica van de Nederlandse zin*. Leuven: Garant.
- Van Haeringen, C.B. 1947. "'Tangconstructies" en reacties daarop'. *De Nieuwe Taalgids*, 40, 1-7.
- Venables, W.N. & Ripley, B.D. 2002. *Modern applied statistics with S*. 4th edn. New York: Springer.
- Verhagen, A. 2005. 'Constructiegrammatica en 'usage based' taalkunde'. *Nederlandse Taalkunde*, 10, 197-222.
- Voorzetsels Prisma – 1999. R. Reinsma, W.J.B. Hus & C. Timmers. Utrecht: Het Spectrum.

- Vrisekoop, M.C. 2001. *Grammaticale termen. Het gereedschap voor elke taal*. Bussum: Coutinho.
- Wasow, T. 1997. 'Transformations and the Lexicon'. In: P. Culicover, T. Wasow and A. Akmajian (eds.). *Formal Syntax*. New York: Academic Press, 327-360.
- Wasow, T. 1997. 'Remarks on grammatical weight'. *Language Variation and Change*, 9, 81-105.
- Wasow, T. 2002. *Postverbal behavior*. Stanford, CA: CSLI Publications.
- Wheeldon, L. & Lahiri, A. 1997. 'Prosodic units in speech production'. *Journal of Memory and Language*, 37, 356-381.
- Wiers, E. & van Noort, M. 1978. 'Woordvolgorde en bijwoordelijke bepalingen.' In: Glot 1978, 193-211.
- Willems, A. & De Sutter, G. 2015. 'Reassessing the effect of the complexity principle on PP placement in Dutch'. *Nederlandse Taalkunde*, 20 (3), 339-366.
- Willems, A. 2016. 'Een extra positie?'. *Nederlandse Taalkunde*, 21 (2), 285-296.
- Wold, S., Esbensen, K. & Geladi, P. 1987. 'Principal Component Analysis'. *Chemometrics and Intelligent Laboratory Systems*, 2, 37-52.
- Yap, M.J. & Balota, D.A. 2009. 'Visual word recognition of multisyllabic words'. *Journal of Memory and Language*, 60, 502-529.
- Zuijdgeest, P. 1995. *Begrijpelijk schrijven. Regels en aanwijzingen voor leesbare, helder geformuleerde officiële brieven en stukken; voorbeelden voor woordkeus en zinsbouw*. Groningen: BoekWerk.
- Zwart, J.W. 1990. 'PP Extraposition from NP in Dutch'. Ms. University of Groningen.
- Zwart, J.W. 1992. Dutch expletives and Small Clause predicate raising. In K. Broderick (ed.). *Proceedings of the Northeast Linguistic Society (Vol. 22)*. Amherst: GLSA, 477-491.
- Zwart, J.W. 1993. Dutch syntax. A minimalist approach. PhD thesis, University of Groningen.
- Zwart, J.W. 2011. *The syntax of Dutch*. Cambridge: Cambridge University Press.

Appendix

Appendix A: lijsten van geraadpleegde algemene literatuur

Appendix A geeft een overzicht van een aantal algemene werken die geraadpleegd werden om een idee te krijgen in hoeverre de variabele positie van VzC's in de Nederlandse hoofd- en bijzin in de taalkundige wereld verankerd is (zie ook §2.1). Achtereenvolgens komen grammatica's en inleidingen in de Nederlandse syntaxis (A.1), woordenboeken van grammaticale termen op het gebied van de Nederlandse taalkunde (A.2) en stijlboeken, taaladvies-, taalverzorgings- en taalzuiveringswerken (A.3) aan bod. Grammatica's en inleidingen in de Nederlandse syntaxis die de woordvolgordevariatie van VzC's vermelden, hebben we gemarkeerd met een asterisk.

A.1 Grammatica's en inleidingen in de Nederlandse syntaxis

- *ANS – *Algemene Nederlandse Spraakkunst*. 1997. Haeseryn, W., Romijn, K., Geerts, G., de Rooij, J. & van den Toorn, M.C. (red.). Groningen/Deurne: Martinus Nijhoff – Wolters Plantyn.
- Balk-Smit Duyzentkunst, F. 1994. *De woorden en hun zin. Grammatica voor iedereen*. Amsterdam: Nijgh & Van Ditmar.
- Balk-Smit Duyzentkunst, F. 2001. *Grammatica van het Nederlands*. Den Haag: Sdu.
- *Bart, P. van, Kerstens, J. & Sturm, A. 1998. *Grammatica van het Nederlands. Een inleiding*. Amsterdam: Amsterdam University Press.
- *Berg, B. van den. 1967. *Beknopte Nederlandse Spraakkunst voor VHMO en kweekschool*. Den Haag: Van Goor.
- Berg, B. van den. 1978. *De grammatica van het Standaardnederlands*. Groningen: Wolters-Noordhoff.
- Calcar, W.I.M. van. 1983. *Een nieuwe grammatica voor taalbeschouwing en taalbeheersing*. Leuven: Acco.
- Cosijn, P.J. & Te Winkel, J. 1888. *Nederlandsche spraakkunst: syntaxis*. Haarlem: Bohn.
- Daenen, C. 1956. *Praktische Nederlandse Spraakkunst: voor de school en voor het leven*. Lier: Van In.
- Droste, F.G. 1965. *Grondbeginselen van de Nederlandse grammatica*. Den Haag: Van Goor.
- *Es, G.A. Van. 1966. *Nederlandse syntaxis in klein bestek*. Zwolle: Tjeenk Willink.
- *Fraussen, E. 1939. *Beknopte Nederlandsche spraakkunst voor middelbaar onderwijs*. Mechelen: St. Franciscus.
- *Groot, A.W. de. 1949. *Structurele syntaxis*. Den Haag: Servire.

- *Hasselbach, W.H. 1905. *Nederlandsche-spraakunst*. Groningen: Wolters.
- *Hertog, C.H. den. 1973. *Nederlandse spraakkunst*. Amsterdam: Versluys.
- Holtvast, K. 1923. *Beknopte Nederlandsche Spraakkunst*. Groningen: Noordhoff.
- Klooster, W.G., Verkuyll, H.J. & Luif, J.H.J. 1969. *Inleiding tot de syntaxis: praktische zinsleer van het Nederlands*. Culemborg/Keulen: Stam-Robijns.
- Klooster, W. 2001. *Grammatica van het hedendaags Nederlands. Een volledig overzicht*. Den Haag: Sdu.
- Kraak, A. & Klooster, W.G. 1968. *Syntaxis*. Culemborg/Keulen: Stam-Kemperman
- Kummer, J.C. 1914. *Nederlandsche spraakkunst*. Amsterdam: S.L. Van Looy.
- Langendonck, J.J.M. Van 1861. *Nederlandsche spraakkunst met oefeningen: voorafgegaan met eenige bedenkingen over eigenbelang, godsdienst en nationaliteit, ten aanzien der taal, en gevolgd van een beknopte woordafleidkunde*. Lier: Van In.
- Ligtenberg, P.H., Van Rossum, H., Schippers, B.W. & Siljée, J.J. 1963. *Nederlands voor de kweekschool. Spraakkunst*. Groningen: Noordhoff.
- Lubbe, H.F.A. Van der. 1978. *Woordvolgorde in het Nederlands. Een synchrone structurele beschouwing*. Assen: Van Gorcum.
- *Luif, J. 1998. *In verband met de zin. Inleiding in de Nederlandse spraakkunst*. Amsterdam: AUP.
- *Olbrechts, G. 1966. *De essentie van het Nederlands*. Antwerpen: De Sikkel.
- *Overdiep, G.S. 1928. *Moderne Nederlandsche grammatica voor gymnasia, HBS en kweekscholen alsmede voor zelfstandige studie*. Zwolle: Tjeenk Willink.
- *Overdiep, G.S. 1949. *Stilistische grammatica van het moderne Nederlandsch*. Zwolle: Tjeenk Willink.
- Overdiep, G.S. & Van Es, G.A. 1956. *Beknopte stilistische grammatica van het Nederlands voor gymnasia, HBS, kweekscholen en lycea*. Zwolle: Tjeenk Willink.
- Paardekooper, P.C. 1960. *Inleiding tot de ABN-syntaksis*. Den Bosch: Malmberg.
- *Paardekooper, P.C. 1963. *Beknopte ABN-syntaksis*. Den Bosch: Malmberg.
- Paardekooper, P.C., Moeyaert, C. & Peperstraete, J. 1963. *Beknopte ABN-spraakunst*. Antwerpen: Scriptoria.
- *Paardekooper, P.C. 1986. *Beknopte Beknopte. Mini-syntaksis van het ABN*. Eindhoven: Eigen beheer.
- Rijpma, E. & Schuringa, F.G. 1925. *Taalverschijnselen. Aanteekeningen en oefeningen bij de Nederlandsche spraakkunst*. Groningen/ Den Haag: Wolters.
- Rijpma, E., Schuringa, F.G. & Naarding, J. 1959. *Nederlandse spraakkunst*. Groningen: Wolters.
- *Rijpma, E., Schuringa, F.G. & van Bakel, J. 1978. *Nederlandse spraakkunst*. Groningen: Wolters-Noordhoff.
- Rijpma, E & Rijpma, J.H. 1956. *Beknopte Nederlandse spraakkunst*. Groningen: Wolters.
- *Schutter, G. De & van Hauwermeiren, P. 1983. *De structuur van het Nederlands: taalbeschouwelijke grammatica*. Malle: De Sikkel.
- *Schutter, G. De. 1974. *De Nederlandse zin. Poging tot beschrijving van zijn structuur*. Brugge: De Tempel.
- *Smedts, W. & Van Belle, W. 2011. *Taalboek Nederlands*. Kalmthout: Pelckmans.
- *Tacx, J.P.M. 1976. *Nederlandse spraakkunst voor iedereen*. Utrecht: Spectrum.
- Terwey, T. & Kuitert, R. 1925. *Nederlandsche spraakkunst*. Groningen: Wolters.
- Tinbergen, D.C. 1932. *Nederlandsche spraakkunst*. Zwolle: Tjeenk Willink.
- *Toorn, M.C. Van den. 1981. *Nederlandse grammatica*. Groningen: Wolters-Noordhoff.
- *Vandeweghe, W. & Devos, M. 2000. *Grammatica van de Nederlandse zin*. Leuven: Garant.
- *Vindevogel, Th. 1958. *Nederlandse grammatica: voor het middelbaar en normaalonderwijs in Vlaams België*. Kapellen: De Sikkel.
- *Vooyo, C.G.N. de & Schönfeld, M. 1967. *Nederlandse spraakkunst*. Groningen: Wolters.

A.2 Woordenboeken van grammaticale termen op het gebied van de Nederlandse taalkunde

- Bart, P. Van & Sturm, A. 1987. *Zinsanalyse en de termen die daarbij gebruikt worden*. Leiden: Nijhoff.
- Pensaert, L. & Van der Sande, L. 1991. *Grammaticaal en taalkundig lexicon: op basis van de Algemene Nederlandse Spraakkunst*. Antwerpen: Standaard.
- Pollmann, T. & Sturm, A. 1977. *Over zinnen gesproken: termen en begrippen van de traditionele grammatica*. Culemborg: Tjeenk Willink/Noorduijn.
- Vrisekoop, M.C. 2001. *Grammaticale termen. Het gereedschap voor elke taal*. Bussum: Coutinho.

A.3 Stijlboeken, taaladvies-, taalverzorgings- en taalzuiveringswerken

- Cockx, P. 1989. *Taalwijzer*. Leuven: Davidsfonds.
- Cockx, P. 1999. *Taalwijzer*. Leuven: Davidsfonds.
- Crombrugge, E. Van, Dooms, L. & Van Nierop, M.J.L. 1973. *Keurig Nederlands in de administratie... en elders*. Gent: Daphne.
- Drop, W. & De Vries, J.H.L. 1977. *Taalbeheersing. Handboek voor taalhantering*. Groningen: Wolters-Noordhoff.
- Gessel, H. van. 2006. *Stijlboek. De Volkskrant*. Den Haag: Sdu.
- Hendrickx, R. 2003. *Stijlboek VRT*. Utrecht: Van Dale Lexicografie.
- Hermkens, H.M. 1974. *Verzorgd Nederlands*. Den Bosch: Malmberg.
- Horst, P.J. van der. 1988. *Taal en tekst van A tot Z. Praktische handleiding voor schriftelijk taalgebruik*. Zutphen: Thieme.
- Horst, P.J. van der. 1993. *Redactiewijzer. Praktische handleiding voor het taalkundig en typografisch verzorgen van teksten*. 's-Gravenhage: Sdu.
- Horst, P.J. van der. 1996. *Stijlwijzer. Praktische handleiding voor leesbaar schrijven*. Den Haag: Sdu.
- Horst, P. van der. 1999. *Redactiewijzer. Praktische handleiding voor het taalkundig en typografisch verzorgen van teksten*. Den Haag/Antwerpen: Sdu/Standaard.
- Klein, M. & Visscher, M. 1992. *Handboek verzorgd Nederlands: spellingregels, stijladviezen*. Groningen: Nijhoff.
- Moerdijk, J.M. & Sluimer, P. 1976. *Taalgids voor de ambtenaar*. 's Gravenhage: VUGA.
- Moerdijk, J.M. & Sluimer, P. 1989. *Taalgids voor de ambtenaar*. Culemborg: Lemma.
- Permentier, L. & Van Den Eynden, L. 1998. *Stijlboek*. Groot-Bijgaarden: Scoop.
- Permentier, L. 2003. *Stijlboek. Onmisbaar voor wie helder wil schrijven*. Roeselare: Roularta Books.
- Renkema, J. 1982. *Schrijfwijzer: handboek voor duidelijk taalgebruik*. 's Gravenhage: Staatsuitgeverij.
- Renkema, J. 1995. *Schrijfwijzer*. 's Gravenhage: Sdu.
- Spek, E. van der. 1996. *Schrijven met perspectief. Structuuradviezen voor schrijvers*. Groningen: Nijhoff.
- Verheij, H. & Kroodsma, E. 1995. *Klinkklare taal. Creatief taalgebruik*. Leuven/Apeldoorn: Garant.
- Verheul, C.F.K. & van Son, J. 1994. *Stijlwijzer*. Apeldoorn: Auctor.
- Verhoeven, G. & Wekking, M. 1991. *Succesvol formuleren. Een advies- en oefenboek voor professionele schrijvers*. Groningen: Wolters-Noordhoff.
- Westen, W. van der. 2002. *Goed geschreven. Zakelijk schrijven binnen opleiding en beroep*. Bussum: Coutinho.
- Witteveen, K. 2006. *Creatief schrijven*. Amsterdam: Boom onderwijs.
- Zuijdgheest, P. 1995. *Begrijpelijk schrijven. Regels en aanwijzingen voor leesbare, helder geformuleerde officiële brieven en stukken; voorbeelden voor woordkeus en zinsbouw*. Groningen: BoekWerk.

Appendix B: voegwoordelijke uitdrukkingen

Voorzetsel + dat	Bijwoord + dat	VzC + dat	Voorzetsel(uitdrukking) + het feit dat
behalve dat buiten dat daargelaten dat doordat mits dat naar dat nadat niettegenstaande dat omdat ondanks dat ongeacht dat opdat sedert dat sinds dat tegen dat tenzij dat totdat uitgenomen dat vanaf dat voordat zonder dat	dermate dat dusdanig ... dat eens dat liever (... dan) dat nu dat of juist dat te zeer ... dan dat vandaar dat zo ... dat zodanig ... dat zodat zodoende dat zolang dat zover(re) dat zozeer ... dat	(al) naargelang dat in afwachting dat in de hoop dat in de overtuiging dat in de veronderstelling dat in de wetenschap dat in die mate dat in die zin dat in geval dat in het besef dat in plaats dat in zover(re) (dat) m.d.v. (met dien verstande dat) met de bedoeling dat met de restrictie dat met de/die beperking dat met het idee dat met het/als gevolg dat met het/als resultaat dat met in het achterhoofd dat onder voorbehoud dat onder voorwaarde dat op een manier dat op het moment dat op een ogenblik dat op het ogenblik dat op voorwaarde dat teneinde dat uit schrik dat uit vrees dat voor het geval dat voor zover(re) dat	aan de hand van het feit dat aangaande het feit dat afgezien van het feit dat als gevolg van het feit dat betreffende het feit dat blijktens het feit dat buiten het feit dat dankzij het feit dat gezien het feit dat in het licht van het feit dat in overeenkomst met het feit dat in overeenstemming met het feit dat in ruil voor het feit dat in tegenstelling met het feit dat in tegenstelling tot het feit dat in verband met het feit dat in weerwil van het feit dat met betrekking tot het feit dat met het oog op het feit dat met uitzondering van het feit dat naar aanleiding van het feit dat naast het feit dat niettegenstaande het feit dat omwille van het feit dat ondanks het feit dat onder invloed van het feit dat ongeacht het feit dat op basis van het feit dat op grond van het feit dat overeenkomstig het feit dat ten aanzien van het feit dat ten gevolge van het feit dat ten opzichte van het feit dat ter ere van het feit dat ter gelegenheid van het feit dat vanwege het feit dat wat betreft het feit dat wegens het feit dat

Samenvatting

In dit proefschrift wordt een relatief onontgonnen type van syntactische variatie onderzocht, meer bepaald de variabele positie van voorzetselconstituenten (VzC's) in de Nederlandse journalistieke bijzin. Zoals uit volgend voorbeeld blijkt, hebben taalgebruikers de keuze om niet-inherente VzC's voor of na de werkwoordelijke eindgroep te plaatsen, i.e. middenveldpositie (zin a) versus achterveldpositie (zin b; ook wel extrapositie):

zin a: [...] dat de trainer **door een laptop naast het veld** vervangen wordt.

zin b: [...] dat de trainer vervangen wordt **door een laptop naast het veld**.

In deze studie leggen we de klemtoon op de variabelen die de distributie tussen de twee woordvolgordevarianten in concreet journalistiek taalgebruik kunnen verklaren. We baseren ons daarvoor op variabelen uit eerdere empirische studies naar dit woordvolgordefenomeen, zoals de complexiteit van het middenveld, de complexiteit van de VzC, de discoursstatus van de VzC en de grammaticale functie van de VzC (Jansen 1978, 1979; Braecke 1990; Hoeksema 2014), en voegen daarnaast nieuwe variabelen toe, zoals de complexiteit van het achterveld, de semantische status van de VzC en de invloed van het werkwoord. Deze studie leest zich dus als een verdere aanvulling, uitbouw en synthese van de bestaande literatuur, met als doel inzicht te geven in de precieze gebruiksvoorwaarden van de twee mogelijke posities van VzC's, die in principe vrij naast elkaar gebruikt kunnen worden.

De reden waarom er op dit moment nog een grote onduidelijkheid heerst over de plaatsingsvoorkeur van VzC's is het monofactoriële karakter van het bestaande empirische onderzoek. Zowel Jansen (1978, 1979) als Braecke (1990) onderzochten het effect van één variabele op de positie van de VzC. Het is dus nog een open vraag hoe de verschillende variabelen zich tot elkaar verhouden in hun effect op de plaatsing van VzC's en of deze variabelen de volledige variatie in plaatsingsgedrag kunnen verklaren. Daarom voeren we in dit proefschrift een logistische-regressieanalyse uit. Met deze multivariate statistische techniek kunnen we heel precies bepalen in welke mate de variabele(n) de keuze tussen middenveld- en achterveldpositie verklaren en voorspellen, rekening

houdend met elkaars effect. Bovendien vindt dit onderzoek zo ook aansluiting bij andere onderzoeken naar syntactische variatiefenomenen in het Nederlands, die wel een statistisch onderbouwde onderzoekstraditie kennen, zoals dat het geval is voor scrambling (van Bergen & de Swart 2010), de datiefalternantie (Colleman 2010), de groene en rode volgorde (De Sutter 2007; Coussé 2008) en de positie van presentatief *er* (Grondelaers & Speelman 2007; Grondelaers, Speelman & Geeraerts 2008; Grondelaers et al. 2009).

Om de repliceerbaarheid en continuïteit in toekomstig onderzoek te garanderen, wat tevens een methodologische doelstelling van dit onderzoek is, moeten de onderzochte zinnen, VzC's en de onderzochte variabelen aan strikte selectiecriteria beantwoorden. Zo focussen we bijvoorbeeld op bijzinnen, om de evidente reden dat de tweede pool in bijzinnen steeds zichtbaar is en er bijgevolg een duidelijk onderscheid is tussen een VzC in het middenveld en een in het achterveld. Omdat het type bijzin een rol kan spelen, beperken we dit onderzoek tot voegwoordzinnen ingeleid door het semantisch lege, grammaticaal onderschikkende voegwoord *dat*.

Voor dit proefschrift gebruiken we enkel data uit de journalistieke component van het Dutch Parallel Corpus (Macken et al. 2011). Door geschreven journalistieke teksten te bestuderen, bieden we een tegengewicht aan de voorgaande empirische onderzoeken die telkens de variabele positie van VzC's in gesproken taalgebruik hebben onderzocht (Jansen 1978, 1979; Braecke 1990). We kunnen ons namelijk afvragen of dezelfde variabelen en mechanismen op eenzelfde manier werkzaam zijn in een geschreven corpus. Anders geformuleerd: het is goed mogelijk dat andere, nieuwe variabelen, die misschien vanuit een gesproken perspectief niet onderzocht werden, een rol spelen in geschreven corpusmateriaal.

De belangrijkste bevindingen van deze studie sommen we puntsgewijs op:

- Achterveldpositie is de meest frequente positie voor VzC's in journalistieke teksten. Dat is een verrassende vaststelling, aangezien zowel traditionele grammatica's als de meer gespecialiseerde literatuur er tot nog toe van uitgingen dat het middenveld de *first choice*-positie is en het achterveld enkel in tweede instantie gekozen wordt (namelijk in het geval dat het middenveld te complex wordt). Daaruit kan men afleiden dat het middenveld de meest frequente positie zou moeten zijn, wat dus niet het geval is voor geschreven journalistiek Belgisch Nederlands. We kunnen deze vaststelling op twee manieren verklaren. Een eerste mogelijke verklaring is corpusgerelateerd. Het is namelijk mogelijk dat de VzC's in dit onderzoek relatief langer zijn dan in het onderzoek naar gesproken taalgebruik. In dat geval is hetzelfde complexiteitsmechanisme werkzaam (nl. korte constituenten voor lange; Dik 1978; Hawkins 1994; Wasow 1997; Stallings, MacDonald & O'Seaghdha 1998), maar met een verschillend resultaat tot gevolg. Een tweede mogelijke verklaring relateert het woordvolgordeverschil aan de

verschillende productieomstandigheden en verwerkingsbehoeften van de spreker en de lezer. We motiveren dat er in een gesproken communicatie sprake is van een spreker-georiënteerd design en in een geschreven communicatie van een lezer-georiënteerd design. Deze psycholinguïstische verklaring moet in vervolgonderzoek op haar waarde getest worden.

- De plaatsing van VzC's is een typisch multifactorieel fenomeen, dat in eerste instantie door de *semantische status van de VzC* bepaald wordt en vervolgens door de *complexiteit*, het *werkwoord* en de *discoursstatus van de VzC*. Dat sommige van die variabelen een invloed hebben op de plaatsing van VzC's wisten we al voordat dit onderzoek uitgevoerd werd, maar met dit onderzoek hebben we een veel vollediger zicht op de variabelen die de plaatsing sturen: we zijn nieuwe variabelen op het spoor gekomen en variabelen waarvan men al lang verwachtte dat ze een effect hebben, zijn weggefallen. We hebben niet alleen een vollediger beeld van de variabelen die de plaatsing van VzC's sturen, we hebben er nu ook een accurater zicht op (in termen van relatieve effecten en *overall*-adequaatheid).
- Er zijn twee interessante interactie-effecten:
 - Het stijgende positieve lineaire effect van de lengte van de VzC op achterveldpositie wordt afgeremd naarmate het achterveld complexer wordt. Deze vaststelling nuanceert het traditionele idee van het achterveld als een (oeverloos) overloopvat.
 - Referenten die in de voorcontext vermeld werden (i.e. toegankelijke referenten) verkiezen in principe middenveldpositie, maar deze tendens wordt afgezwakt naarmate het informationele belang van de referent toeneemt. Omgekeerd verkiezen ontoegankelijke of nieuwe referenten in principe achterveldpositie, maar deze voorkeur wordt afgeremd als de referent veeleer onbelangrijk is.
- Rekening houdend met het multifactoriële karakter van de woordvolgordevariatie gaan we ervan uit dat verschillende mechanismen simultaan werkzaam zijn en dat er bijgevolg verschillende verklaringsprincipes simultaan geldig kunnen zijn. Volgens ons wijzen de significante predictoren in de richting van een functioneel-cognitieve verklaring. Een interessante vraag voor vervolgonderzoek is of ook andere verklaringsmechanismen een rol kunnen spelen.

Summary

This doctoral thesis studies a relatively unexplored type of syntactic variation, in particular the positional variations of prepositional phrases (PPs) in Dutch subordinate clauses. Non-inherent PPs can be placed before or after the final verb cluster, i.e. they can go in middle-field position (sentence a) or in post-field position (also known as extraposition; sentence b) as illustrated in the following example:

sentence a: [...] dat de trainer **door een laptop naast het veld** vervangen wordt.

[...] that the coach **by a laptop at the side of the field** is replaced

sentence b: [...] dat de trainer vervangen wordt **door een laptop naast het veld**.

[...] that the coach is replaced **by a laptop at the side of the field**

The present study focuses on the variables which determine both variations in word order in journalistic language. It uses the variables established in previous empirical studies of positional variation of PPs as a starting point, including the *complexity of the middle-field*, the *complexity of PPs*, the *discourse status of PPs* and the *grammatical function of PPs* (Jansen 1978, 1979; Braecke 1990; Hoeksema 2014). Several new variables are added to these, such as the *complexity of the post-field*, the *semantic status of PPs* and the *verb*. The literature on the subject is summarized, expanded and developed further in order to specify the conditions governing usage of the two positional variations of PPs, which are basically in free variation.

In previous research, the lack of clarity about the conditions governing the preferred position of PPs was due to the monofactorial statistical approach. Both Jansen (1978, 1979) and Braecke (1990) analyzed the effect of a single variable on the position of PPs. So it remains to be determined how the different variables interact, how this affects the position of PPs and whether the selected variables explain the positions of PPs in texts. To answer these questions, we use a logistic regression analysis. This multivariate statistical technique allows us to determine exactly to what degree the different variables explain and predict the choice between the middle-field and post-field positions, taking into account how they affect each other. Furthermore, the present study ties in with research into other instances of syntactic variation in Dutch, founded on statistical

analysis. These include scrambling (van Bergen & de Swart 2010), dative alternation (Colleman 2010), the so-called green and red word order in verb clusters (De Sutter 2007; Coussé 2008) and the position of presentative *er* “there” (Grondelaers & Speelman 2007; Grondelaers, Speelman & Geeraerts 2008; Grondelaers et al. 2009).

One of the methodological aims of this study is to ensure the reproducibility of the findings and their value for further research. To achieve this the sentences and PPs in the corpus and the variables examined were selected subject to strict criteria. Our research focuses on subordinate clauses, as these always contain a second pole, which offers a clear distinction between PPs in the middle-field and PPs in the post-field. The type of subclause may be a significant factor; this is why we selected subclauses introduced by the subordinating conjunction *dat*.

This study uses data taken exclusively from the journalistic texts in the Dutch Parallel Corpus (Macken et al. 2011). Using written journalistic texts allows us to depart from the spoken corpus-based previous empirical studies on PPs (Jansen 1978, 1979; Braecke 1990). Different variables and mechanisms may operate in spoken corpus data and written data. We therefore investigate the role of established variables in written texts and add new ones which may not have been relevant in spoken language.

The main findings of this study are the following:

- Post-field position is the most frequently used position for PPs in Belgian journalistic texts. This is surprising, as both traditional grammars and specialist literature have always assumed that the middle-field is the first choice position and that the post-field is merely a second option (in case of a complex middle-field). This proposition can be explained in two ways. The first explanation is corpus related. The PPs in our written corpus may be longer compared to those in the spoken corpus. If this is the case, the same complexity mechanism applies but generates different findings (viz. short phrases precede long phrases; Dik 1978; Hawkins 1994; Wasow 1997; Stallings, MacDonald & O’Seaghdha 1998). The second explanation suggests that the difference in word order is linked to differences in the circumstances in which the text is produced and in the way the reader and the speaker process the text. We argue that spoken communication is speaker oriented while written communication is reader oriented. Further psycholinguistic research will have to test the validity of this explanation.
- The position of PPs is a typically multifactorial phenomenon, which is determined in hierarchical order by the *semantic status of the PP*, the *complexity*, the *verb* and the *discourse status of the PP*. The impact of some of these variables was known at the beginning of this study, however, we now have a much more comprehensive idea of the part these variables play in this type of syntactic word order variation. We found new variables that have an effect and other variables which have long been suspected to have an effect, proved not to have one. This means that we now not

only have a better idea of which variables determine the position of PPs, but we also have a more accurate model (in terms of relative effects and overall adequacy).

- There are two interesting interaction effects:
 - We found an interesting interaction effect of the *complexity of the PP* weight and the *complexity of the post-field*, which confirms the importance of the PP complexity in PP placement, with the post-field complexity as an inhibitory variable: the heavier the post-field, the fewer PPs in post-field position. This finding puts the traditional view of the post-field as an expansion tank in perspective.
 - The preferred position for referents which are accessible (because they are part of the context before the sentence in question) is generally middle-field position. However, this principle applies less strongly when the information importance of the referent increases. Less accessible or new referents, on the other hand, tend to occur more often in post-field position, although this principle applies less strongly when the referent is rather unimportant.
- Taking into account the multifactorial nature of this word order variation, we assume that different mechanisms work simultaneously and that different explanatory principles are simultaneously active as a result. In our opinion, the significant predictors of our model point towards a functional cognitive explanation. An interesting question for further research is whether there are other mechanisms that can explain this type of word order variation.